CRISPR-CAS9 BASED GENETIC ENGINEERING AND MUTATION DETECTION IN GENUS *NICOTIANA*

Tuukka Huhdanmäki
Master's Thesis
University of Helsinki
Department of
Agricultural Sciences
Plant Breeding
2021

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Maatalous-metsätieteellinen tiedekunta	Maataloustieteiden laitos

Tekijä — Författare — Author
Tuukka Huhdanmäki

Työn nimi — Arbetets titel — Title
CRISPR-Cas9 based genetic engineering and mutation detection in genus *Nicotiana*

Oppiaine —Läroämne — Subject
Maataloustieteet, Kasvinjalostus

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Maisterintutkielma	Huhtikuu 2021	86 s.

Tiivistelmä — Referat — Abstract

CRISPR-Cas9-menetelmä on yksi uusista geeniteknologian läpimurroista, joka hyödyntää kohdennettua mutageneesiä. Menetelmä perustuu Cas-perheen proteiineihin ja opas-RNA:han, jotka yhdistettynä mahdollistavat kohdentamisen valittuun sekvenssiin ja sen muokkaamisen halutulla tavalla. Tämän maisterintutkielman tavoitteena oli verrata eri CRISPR-Cas9-menetelmään perustuvia geenimuokkaustekniikoita *Nicotiana*-suvun kasveissa sekä onnistuneen mutaation tunnistamiseen tarvittavia tekniikoita. *PDS1* ja *PDS2* ovat geenejä, jotka koodaavat fytoeeni desaturaasi-entsyymin toimintaa kasveissa. Nämä geenit valittiin kohdegeeneiksi, koska mutatoitunut geeni saa aikaan valoherkän fenotyypin, joka on helppo tunnistaa viherhiukkasten vaurioitumisesta johtuvasta valkolaikukkaasta ilmiasusta.

CRISPR-Cas9 ribonukleoproteiinikompleksivälitteinen transformaatio (RNP) käyttää hyväkseen erikseen tuotettua Cas9-proteiinia ja opas-RNA:ta, jotka yhdistettynä aikaansaavat ohimenevän geenimuutoksen solussa. Tätä tekniikkaa oli tarkoitus käyttää, mutta liukoisessa muodossa olevan Cas9-proteiinin tuottaminen koitui ongelmalliseksi eikä lopullista transformaatiota päästy tekemään. Tämän tutkimuksen valossa valmiin Cas9-proteiinin hankkiminen saattaa olla kannattavampaa kuin proteiinin tuottaminen alusta alkaen, kun tavoitteena on aikaansaada vain muutamia geenimuunneltuja kasveja.

Agrotransformaatio on *Nicotiana*-suvulla hyväksi todettu agrobakteereita hyödyntävä geeninsiirtotekniikka ja kun tähän yhdistetään STU-CRISPR-Cas9-menetelmä, saatiin aikaan agrobakteerikantoja, jotka suorittavat geenimuokkauksen alusta loppuun yhden sekä Cas9-proteiinin että opas-RNA:n tuottavan plasmidin avulla. Onnistuneesti muokattuja kasveja saatiin aikaan sekä agroinfiltraatio että agrotransformaatio tekniikoilla. Menetelmän heikkoutena ovat mahdolliset off-target -mutaatiot, joita saattaa olla tarve ristisiittää pois populaatiosta. Tämä menetelmä ei sovellu kaikille eliöille ja on rajallisempi kuin RNP-välitteinen transformaatio, mutta on toiminnaltaan yksinkertainen ja soveltuva *Nicotiana*-sukuun.

Onnistuneet mutaatiot tunnistettiin käyttämällä kaupallista T7E1 endonukleaasia sekä luonnollista CEL I endonukleaasia sisältävää selleriuutetta, jonka havaittiin soveltuvan edullisena vaihtoehtona tulosten varmistamiseen ja toistamiseen.

Avainsanat — Nyckelord — Keywords
CRISPR-Cas9, Transformaatio, *Nicotiana*, Ribonukleoproteiini, sgRNA*, PDS1*, *PDS2*

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Prof. Teemu Teeri

Säilytyspaikka — Förvaringsställe — Where deposited
Maataloustieteiden laitos ja Viikin kampuskirjasto

Muita tietoja — Övriga uppgifter — Further information

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Agriculture and Forestry	Department of Agricultural Sciences

Tekijä — Författare — Author
Tuukka Huhdanmäki

Työn nimi — Arbetets titel — Title
CRISPR-Cas9 based genetic engineering and mutation detection in genus *Nicotiana*

Oppiaine —Läroämne — Subject
Agricultural Sciences, Plant Breeding

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	April 2021	86 p.

Tiivistelmä — Referat — Abstract

CRISPR-Cas9 is one variant of newly emerging technologies utilizing targeted mutagenesis based on Cas family proteins and guide RNA that enable binding and modifying selected target sequence. The aim of the master's thesis was to compare different methods of CRISPR-Cas9 induced gene editing in the genus *Nicotiana* and other secondary protocols necessary to identify successful mutations. *PDS1* and *PDS2* genes coding phytoene desaturase in plants were selected as target genes as mutant genotype produce visually identifiable photobleaching phenotype.

CRISPR-Cas9 ribonucleoprotein complex mediated transformation uses separately produced Cas9 protein and guide RNA that when combined perform transient gene editing in cell. This method was planned to be used but Cas9 protein was challenging to produce in soluble form and final transformation was not achieved. This study suggests that acquiring ready-to-use Cas9 protein might be preferable choice when targeting only few transformations with CRISPR-Cas9 RNP-complex.

Agrotransformation is well established method for genus *Nicotiana* and using Single Transcriptional Unit CRISPR-Cas9 system it is straightforward procedure from plasmid design to transformation. Successfully transformed plants were redeemed from transient agroinfiltration and stable agrotransformation experiments. Off-target mutations are possible and selective outbreeding may be needed. This method lacks the several advantages of CRISPR-Cas9 RNP-complex such as instant gene editing in cell, avoiding RNA interference and transformation over species boundaries, but is simple and functional in genus *Nicotiana*.

Successful mutations were detected using commercial T7E1 and with natural CEL I endonuclease from celery extract. Celery extract can be used as cost-effective alternative to T7E1 for verifying or replicating previously confirmed results.

Avainsanat — Nyckelord — Keywords
CRISPR-Cas9, Transformation, *Nicotiana*, Ribonucleoprotein, sgRNA, *PDS1*, *PDS2*

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Prof. Teemu Teeri

Säilytyspaikka — Förvaringsställe — Where deposited
Department of Agricultural Sciences and Viikki Campus Library

Muita tietoja — Övriga uppgifter — Further information

Contents

ABBREVIATIONS AND CONSEPTS

Cas9	CRISPR associated protein 9
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
dCas9	Nuclease-deficient (dead) Cas9
DNA	Deoxyribonucleic acid
DSB	Double-stand break
gRNA	Guide RNA
NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
PDS1	Phytoene desaturase 1
PDS2	Phytoene desaturase 2
pegRNA	Prime editing guide RNA
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RVD	Repeat variable diresidues
sgRNA	Single guide RNA
siRNA	Small interfering RNA
SNP	Single-nucleotide polymorphism
STU	Single transcriptional unit
TALEN	Transcription activator-like effector nuclease
tracrRNA	Trans-activating CRISPR RNA
ZFN	Zinc finger nuclease

1 INTRODUCTION

Emergence of genetic engineering in the seventies revolutionized wide spectrum of natural sciences changing the way how humanity perceives surrounding natural world. The concept of hereditary information was realized by Gregor Mendel in the late nineteenth century and refined into understanding of genetic information in the early twentieth century, but it was not before advent of gene transfer that mechanistical nature of life was uplifted as an ideology and full potential of biotechnology would be unraveled (Jackson et al. 1972, Cohen and Chang 1973). The first wave of genetically engineered crop plants experienced strong resistance through ideological fear and suspicion towards new technology seemingly diverting from 'natural' but also suffered from hype letdown caused by inferior products marketed in sublime ways for consumers (Hiatt 1994, Losey et al. 1999).

While consumers and politicians continued to refuse potentially enhanced products, the technology behind genetic engineering evolved in the hands of devoted scientists. New technology exploiting Zinc finger nucleases (ZFNs) (Bibikova et al. 2003) and Transcription activator-like effector nucleases (TALENs) (Boch et al. 2009) emerged in the first decade of the new millennia bringing much desired accuracy into gene transfer. Instead of adhering randomly into indefinite part of genome, a specific part of genome or genes could now be targeted with precise accuracy and changed systematically. The newly founded programmable nucleases were however arduous to engineer and targeted modification of entire gene was still challenging.

A breakthrough came in 2012 when Charpentier and Doudna discovered and modified CRISPR-Cas9 protein complex in *Streptococcus* that protects bacterial cells from virus by identifying and obliterating viral DNA (Jinek et al. 2012). Harnessed with a manmade guide RNA this complex could now be directed into specific genes with superior accuracy and genes could now be altered from single nucleotide changes all the way to adding or replacing entire genes. Future possibilities are virtually unlimited, which allows great scientific progress to be made and commercialized for consumers, possibly enhancing life quality in global scale. Unlimited possibilities come with great concern in form of using this technology with malicious intent. That however should not diminish CRISPR-Cas9 research as the more is known the more can be done to help preventing misuse and to direct resources for new revolutionary inventions.

2 LITERATURE REVIEW

2.1 Early days of targeted mutagenesis

The history of genetic engineering in plants begins from first *Agrobacterium* based gene transfer method developed in 1983 by Shaw and colleagues when their team successfully transformed tobacco seedlings with *Agrobacterium* strain containing tumor inducing plasmid with inserted rabbit β-globin gene (Shaw et al. 1983). This method became prevalent technique to introduce stable gene modifications into plants while second method developed in 1987 by Klein and colleagues based on high-speed particle bombardment using microscopic tungsten or gold particles coated with DNA prepared for transfer, has since been used mainly in gene expression research due to its low efficiency to produce stable transformants (Klein et al. 1987).

Disconnected genetic material, either endogenic or inserted, is integrated into stable part of genome by two primary modes. These are non-homologous end joining (NHEJ) in which gene insert is ligated directly into DNA strand containing double-stand break (DBS) without homologous template and homologous recombination where break is repaired and ligated based on existing template (Pfeiffer et al. 1994, Moore and Haber 1996). NHEJ as primary integration method in plants involving specific proteins was demonstrated in *Agrobacterium* based gene insertion by Van Attikum and colleagues in 2001, while biolistic delivery was studied by Liu and colleagues in 2019 and NHEJ was found to be primary integration method also in gene insertion based on mechanical delivery where cell's endogenous processes integrate inserted DNA into DSBs generated by mechanical force (van Attikum et al. 2001, Liu et al. 2019). Both techniques cause random integration where one or multiple copies of inserted gene is attached to any available DSB in any possible chromosome. This can lead to disruption of other genes or cause unstable production of desired gene product if excessive amount of copies is added.

Accuracy was desperately needed in genetic engineering and first techniques adding precision into gene delivery arose in form of zinc finger nucleases, Transcription Activator-like Effector Nucleases (TALEN) and meganucleases. These nucleases function through the same basic principle of target sequence binding and cutting leading to a double-strand break, but vary in their chemical structure and sequence specificity.

First zinc finger structure was originally found from model frog species (*Xenopus laevis* Daudin, 1802) in which it occurs in specific transcription factor TFIIIA enhancing its binding strength (Klug and Rhodes 1987). Construction process of human engineered ZFNs was introduced in 2003 by Bibikova and colleagues when their team engineered Zinc Finger Nuclease targeting X chromosomal *yellow* (y) gene affecting melanin production in the common fruit fly *Drosophila melanogaster* (Meigen, 1830) (Bibikova et al. 2003). Zinc finger structures appear in zinc finger nucleases (ZFN) where they function as pair of domains attached to opposite DNA strands and bound around desired cutting site. Each recognition domain is composed of three linked Cys_2His_2 zinc fingers in which individual zinc finger recognizes three adjacent base pairs (bp) thus giving single domain recognition capacity of 9 bp while the entire ZFN complex recognizes continuous sequence of 18 bp (Kim et al. 1996, Segal et al. 1999). This is significant length for recognition site that can be only 4-8 bp long and still deliver precise enough recognition for many natural restriction enzymes. See Figure 1A for a schematic diagram of ZFN complex binding to its target site.

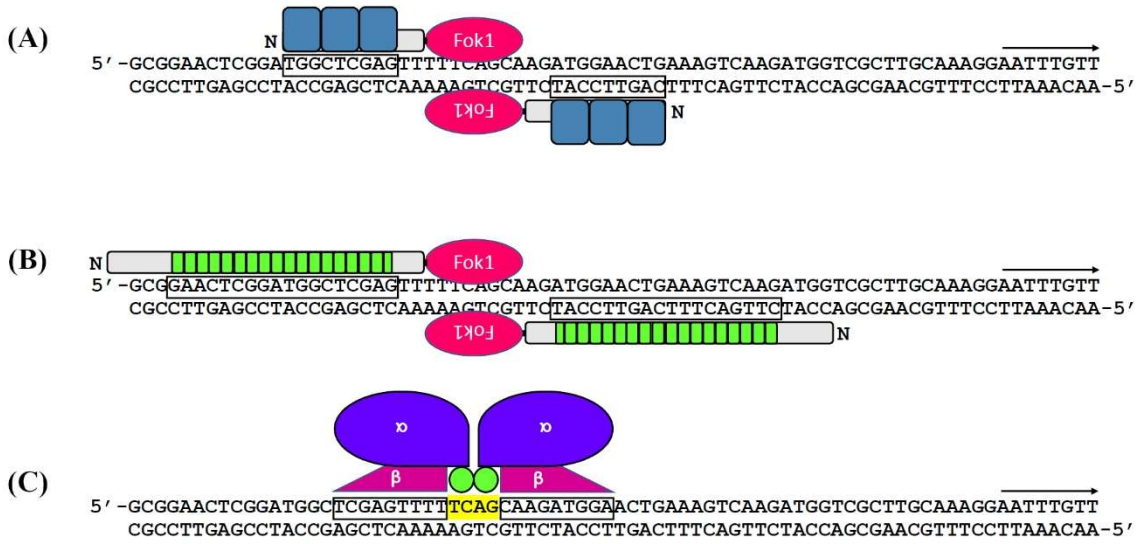


Figure 1. Different nucleases and their DNA targeting mechanism illustrated. **(A)** ZFN with opposite recognition domains in blue. Each division represents single zinc finger protein. The FokI nuclease is represented as a red oval. Arrow represents sequence reading direction. **(B)** TALEN with opposite recognition domains in green. Each division represents single repeat module. **(C)** Meganuclease homodimer with enzyme segments containing α-helices in blue and sequence recognizing β-ribbons in plum color. Green dots represent the active site with two bivalent metal ions that have activating and inhibiting functions.

TALEN mechanism was described from a plant pathogen causing bacterial leaf spot (*Xanthomonas campestris pv. vesicatoria* (Doidge) Dye) in which it is found as AvrBs3 transcription activator-like effector (TALE) that causes hypertrophic growth of mesophyll cells (Kay et al. 2007). Similar to ZFNs, when two TALEs are attached into nuclease they form transcription activator-like effector nuclease (TALEN) where they function as central domain binding around target site in the opposite DNA strands but unlike in triplet based ZFNs, TALEN recognition is based on tandem repeats each consisting of two interlinked helix-structures with hypervariable amino acid residues in their conjoined tip, which in this context are known as repeat variable diresidues (RVDs) (Deng et al. 2012, Mak et al. 2012). These two RVDs jointly recognize a single base pair in the target DNA, thus recognition capacity of the entire TALEN molecule is directly determined by number of tandem repeats. The first described natural TALE AvrBs3 consists of 17.5 tandem repeats per domain thus giving the entire molecule functional recognition capacity of 34 base pairs in the target sequence (Moscou and Bogdanove 2009, Boch et al. 2009). See Figure 1B for a schematic diagram of TALEN complex binding to its target site.

TALEN engineering was pioneered in 2011 by Miller and colleagues when their team produced a TALEN with cutting capabilities by adding the catalytic domain of FokI restriction endonuclease into the molecule structure (Miller et al. 2011). First actual gene modifications with engineered TALENs were done by two individual teams on zebrafish (*Danio rerio* F. Hamilton, 1822) later in 2011 with one team targeting BamHI cutting site of kinase gene *tnikb* (Huang et al. 2011) and other team targeting *gria3a* (AMPA glutamate receptor) and *hey2* (transcription factor) genes (Sander et al. 2011).

Meganucleases are a group of homing endonucleases originally found from yeast (*Saccharomyces cerevisiae* Meyen ex E.C. Hansen) in the early eighties, first represented by a type affecting mating-type switching in yeast (Kostriken et al. 1983) and a second type responsible for intron splicing in mitochondrial DNA in yeast (Jacquier and Dujon 1985). After these initial findings hundreds of other meganucleases (Grizot et al. 2010) have been found, from which most studied ones belong to the LAGLIDADG family of meganucleases sharing homologous nonapeptide motifs abbreviated as LAGLI and DADG by their amino acids in one-letter codes (Hensgens et al. 1983).

Structural mechanism of target site recognition and cleaving was first described from I-CreI meganuclease in 1998 by Jurica and colleagues. Meganucleases are formed from two monomeric enzymes combined into a homodimer by two α-helices containing the LAGLIDADG motif. This axis contains the active site performing sequence cutting and divides the meganuclease into two apparent domains and causes the characteristic ribbon-like structure for meganucleases. The active site contains two calcium ions that inhibit the cutting reaction and are substituted by either magnesium or manganese ions, depending on the specific meganuclease, when cutting reaction is assigned. Target-site recognition is performed by the β-ribbon that is a sheet-like structure under the domains attaching to the target sequence. Different meganucleases have varying sequence recognition capacity, but in the case of I-CreI, length of recognized homing site is 22 base pairs (Jurica et al. 1998). See Figure 1C for a schematic diagram of meganuclease complex binding to its target site.

First experiments with meganuclease engineering were done in 2003 by Epinat et. al where their team designed two types of meganucleases of which the first, based on I-CreI, could only induce homologous recombination in cells comparable to its unmodified parent protein as only its chemical structure was altered to form single chain but the second meganuclease was engineered by combining monomers from two different natural meganucleases I-CreI and I-DmoI from the LAGLIDADG family into a new hybrid protein that they named DmoCre. This new protein had a synthetic target site differing from its parent proteins (Epinat et al. 2003).

Due to the complex structure and challenging engineering of meganucleases their use has been concentrated on basic research and cell level homologous recombination, and only few experiments have so far produced organisms with stable and heritable gene modifications. The first plant modified with engineered meganuclease based on I-CreI was maize, introduced in experiments by Gao and colleagues in 2010. Their team targeted the *LIGULELESS1* gene that affects formation of ligules and auricles in maize. The meganuclease construct was delivered into immature embryos by *Agrobacterium*-mediated transformation and they managed to produce plants with short deletions in the gene that were also inherited by their progeny even though deletions were not severe enough to produce plants with liguleless phenotype. This indicates the deletions could have been monoallelic (Gao et al. 2010).

2.2 Introduction to the CRISPR-Cas9 system

2.2.1 Origin of CRISPR-Cas9

History of CRISPR-Cas9 system began in 1987 when team led by Yoshizumi Ishino was studying *iap* gene in *E. coli* and accidentally found unusual fivefold repeating sequence in the genes 3' flanking end that was constructed from 29 nucleotides long homologous repeats separated by 32 nucleotides long spacing (Ishino et al. 1987). This structural class was later named as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) (Jansen et al. 2002). Same paper by Jansen and colleagues established name "CRISPR associated" genes and named four Cas genes (*Cas1, Cas2, Cas3* and *Cas4*) after the genes they found bordering the CRISPR loci in the prokaryotic species examined in the study.

Purpose of these repeats started to unveil around 2005 when three different groups identified connection between CRISPR sequences in microbes and corresponding sequences in different bacteriophages infecting these microbes and drew conclusions that this connection was related to bacteriophage immunity in microbes (Bolotin et al. 2005, Mojica et al. 2005, Pourcel et al. 2005). Mechanism of action was fully explored in 2007 by Barrangou and colleagues when their team studied bacteriophage resistance in *Streptococcus thermophilus* (Orla-Jensen 1919) and discovered the cumulative nature of this resistance where after viral infection the prokaryote could integrate new spacer elements originating from the infecting bacteriophage genome into its own genome by using Cas gene encoded enzymes (Barrangou et al. 2007). The prokaryote can later use these parts of viral genome called protospacers to produce RNA sequences to be used in RNA interference-based defense against the previously met bacteriophage (Makarova et al. 2006).

Breakthrough for harnessing this natural resistance system into research tool to be utilized by man came in 2012 when group led by Jennifer Doudna and Emmanuelle Charpentier, who now take credit for the invention, published a study where they and their teams showed for the first time that CRISPR-Cas system functions actively by producing double-stranded breaks into target DNA with a Cas nuclease that is guided to the target sequence by CRISPR RNA-guide sequences (crRNA). The crRNA can be engineered to

target any part of genome, thus providing unprecedented accuracy and ease of design into experimenting in the field of genetic engineering (Jinek et al. 2012).

2.2.2 CRISPR-Cas9 mode of action

In the experiments from 2012 by Doudna and Charpentier selected as their study subject the Cas9 nuclease based CRISPR-Cas system (CRISPR-Cas9), seeing it as the simplest in function compared to systems based on other nucleases from the Cas protein family. In the natural unmodified CRISPR-Cas9 system sequence recognition is based on two interlocking RNA molecules, produced separately and merged in the completed CRISPR-Cas9 system. The first RNA, called CRISPR RNA (crRNA), contains a complementary sequence attaching to the target DNA while the second RNA, called trans-activating crRNA (tracrRNA), is attached to the free end of crRNA by base pairing with the crRNA. The complex recognizes only target sequences that are followed by protospacer adjacent motif (PAM) sequence present in the targeted viral genome, which in case of Cas9 is simply three letter sequence of NGG where N marks any nucleotide followed by two guanines. PAM site is needed for cutting site recognition and when recognized by crRNA and activated by tracrRNA, Cas9 nuclease will cut both DNA strands three base pairs upstream from the PAM sequence (Jinek et al. 2012).

Regarding genetic engineering, a system based on two different RNAs where both are essential for function is unpractically complicated. Combining these RNAs into joint complex was achieved by Doudna's and Charpentier's teams by fusing the 3' end of crRNA into the 5' end of the tracrRNA with hairpin-like linker loop, forming a single guide RNA (sgRNA) that mimics the activated natural dual-complex but can be produced entirely in single gene transcription event. In addition to the fusion, the structure was simplified by cutting off excess sequence from the 3' end of tracrRNA. It was discovered that for successful recognition and activation of Cas9, only a 5 to 12 nucleotides long tail following the base pairing interaction between crRNA and tracrRNA was needed (Jinek et al. 2012). See Figure 2 for a schematic diagram of engineered CRISPR-Cas9 complex binding to its target site.

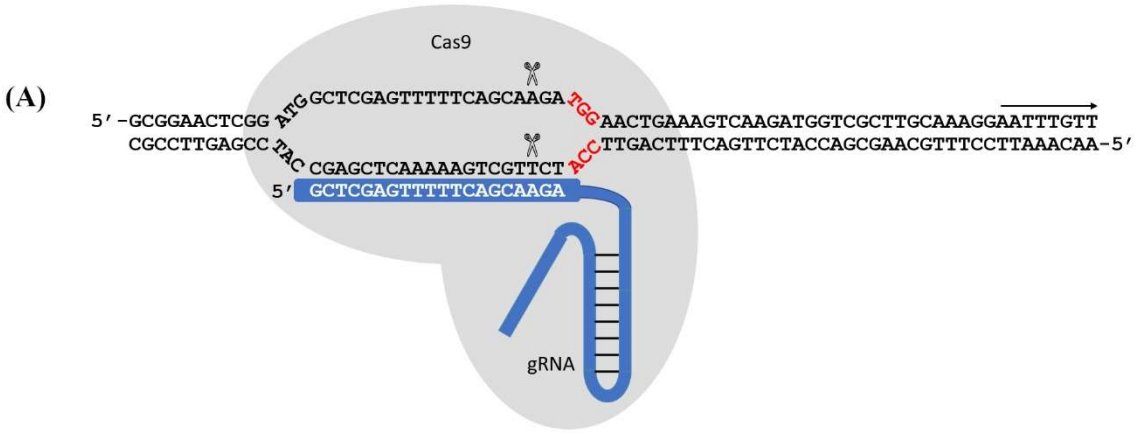


Figure 2. CRISPR-Cas9 mode of action illustrated. The Cas9 enzyme attached to target sequence is illustrated by grey shading. Sequence recognizing gRNA in blue demonstrating attachment to single strand of unwind DNA helix. Scissor symbols represent cutting sites next to the PAM sequence marked in red. Arrow represents sequence reading direction.

2.3 Ribonucleoprotein complex mediated CRISPR-Cas9 system

Ribonucleoproteins (RNPs) are protein complexes containing RNA. Ribonucleoprotein mediated CRISPR-Cas9 gene editing means a technique where the Cas9 protein and the guide RNA are preassembled outside the cells and later transferred into cells as a functioning complex that performs transient gene editing. The method was first described in 2013 by Cho and colleagues when their team performed gene editing in the nematode *Caenorhabditis elegans* (Maupas, 1900) by direct injection of Cas9–sgRNA ribonucleoproteins. Compared to plasmid-based transformation where an entire plasmid containing instructions to translate Cas9 protein and guide RNA is introduced into cell, the ribonucleoprotein-based method has several advantages as it performs gene editing instantly without the need to first produce components from the plasmid before editing can occur. Ribonucleoprotein complexes cannot be silenced by the cell unlike exogenous messenger RNA from plasmids but will degrade quickly after the desired gene editing has been performed. RNP complexes can be used over species boundaries due to no need for adjusting organism specific promoters that are needed when plasmids are used (Cho et al. 2013).

New delivery methods of RNP complexes into cells were later described by other teams. Electroporation-based delivery method with nucleofector technology into human cells was described by Kim and colleagues in a study from 2014 where they managed to

perform gene editing in great efficiency with up to 79% of cells gaining site-specific mutations while also observing a lower amount of off-target mutations due to rapid degradation of RNPs after delivery compered to plasmid-based methods (Kim et al. 2014). Delivery methods to plants were developed by Woo and colleagues in 2015 when their team edited protoplasts of several species via polyethylene glycol (PEG)-mediated transformation, while biolistic delivery was developed in 2016 by Zhang and colleagues to wheat immature embryo cells and in 2016 by Svitashev and colleagues to maize embryo cells (Woo et al. 2015, Zhang et al. 2016, Svitashev et al. 2016).

2.4 *Agrobacterium*-mediated CRISPR-Cas9 system

First *Agrobacterium*-mediated gene modifications with CRISPR-Cas9 system in *Nicotiana benthamiana* (Domin) and *Nicotiana tabacum* (Linné) were made with vector constructs containing separate *Cas9* gene sequence and sgRNA expression cassette (Li et al. 2013, Nekrasov et al. 2013, Gao et al. 2015). While this two-part system works well in most cases it is not optimal in all situations. In 2016 Tang and colleagues introduced a single transcriptional unit (STU) CRISPR-Cas9 system for genetic engineering in plants in which single Pol II promoter operates expression of both Cas9 protein and sgRNA (Tang et al. 2016). This system is based on self-cleaving hammerhead ribozyme introduced by Haseloff and Gerlach in 1988 that when added to this system cuts sgRNA from the full mRNA transcript (Haseloff and Gerlach 1988). Full mRNA transcript before translation contains Cas9 followed by terminating synthetic polyA site followed by first hammerhead ribozyme cutting site before the sgRNA and a second cutting site before ending in the actual hammerhead ribozyme motif. Translation process starts with production of Cas9 and ends in a synthetic polyA while self-cleaved Hammerhead ribozyme cuts off the sgRNA part of mRNA (Tang et al. 2016).

Advantages of STU CRISPR-Cas9 system include fully controlled expression of Cas9 and sgRNA that allows regulated induction, easy manipulation of multiple sgRNAs and single promoter that eliminates potential compatibility problems arising from use of other promoters (Tang et al. 2016).

2.5 gRNA design

Guide RNA (gRNA) is a collective term for both crRNA:tracrRNA complex and sgRNA based methods when target sequence designing is discussed in general. gRNA designing is essentially picking up an appropriate target-site in gene of interest and depending on the used delivery method and targeted organism, choosing suitable promoter that functions in sync with the other components necessary to perform successful gene editing in the target cell. When gRNA and tracrRNA are used in unison instead of a sgRNA containing hybridized version of both, these are designed and produced individually. gRNA target-site is twenty nucleotides long sequence that is chosen from part of gene so that it is directly followed by the PAM sequence NGG. Guanine next to PAM site and sgRNA containing more than 50% of guanine and cytosine were found to be factors explicitly enhancing cleavage rate in target cells (Gagnon et al. 2014).

The specific part of the sequence to be targeted is chosen depending on desired outcome whether it is frameshift mutation by deletion or insertion of base pairs or insertion of larger gene segment altering gene function. Risk analysis for off-target mutations should be done especially when humans or other multicellular animals are targeted. In plants individuals with off-target mutations are easier to manage as they can be removed with selection or the mutation can be outbred by crossbreeding. Possible off-targets can be identified simply by using online sequence aligning tool where desired target sequence is compared against full genome of the organism and similar or identical sequences are displayed by the program. These matching sequences are probable off-target mutation sites if they are next to a suitable PAM site.

2.5.1 gRNA design in RNP system

In a study from 2014 by Gagnon and colleagues efficiency of two promoters T7 (5'-TAATACGACTCACTATA-3') and SP6 (5'-ATTTAGGTGACACTATA-3') was compared and SP6 was found to be superior of the two as in vitro transcription with T7 contained errors resulting from initiation site related confusion caused by T7 RNA polymerase attaching to other bases than actual 5'GG initiation site (Gagnon et al. 2014).

Gene specific oligomer is designed as following sequence:

P-**N_{20}**-GTTTTAGAGCTAGAAATAGCAAG

Sequence starts with promoter (P) followed by **N_{20}** that is the selected target sequence and ending in overlapping region that anneals to constant oligonucleotide (tracrRNA), which in protocol by Gagnon and colleagues, is produced separately.

2.5.2 sgRNA design in STU system

Single transcriptional unit (STU) CRISPR-Cas9 system produces both Cas9 and sgRNA from the same transcript. When mediated by *Agrobacterium*, the sgRNA is designed into the vector sequence instead of synthetically produced oligomer as in RNP system. In 2016 Tang and colleagues produced ready-made STU expression vectors where the twenty base pair long target sequence is simply inserted into the vector's empty scaffold section between hammerhead ribozyme cleavage sites with Golden Gate cloning method (Tang et al. 2016).

2.6 Future of CRISPR-Cas technologies

2.6.1 Versatile editing with new CRISPR-Cas systems

Necessity of PAM sequence NGG when using Cas9 considerably hinders possible target sequences in any given gene of interest when direct sequence editing is desired. This strict rule has started to crumble in past years as several other Cas proteins with alternative PAM sites have been studied and new varieties of Cas9 protein have been developed that enable new alternative PAM recognition sites making target site selection more versatile. In 2018 two different Cas9 varieties were developed, first one called xCas9 was developed by Hu and colleagues with phage-assisted continuous evolution to mutate wild type Cas9 to accept NG, GAA and GAT nucleotide combinations as PAM sites (Hu et al. 2018). Second new variety called SpCas9-NG was developed by Nishimasu and colleagues in 2018 and later additionally studied by Ren and colleagues in 2019 and was confirmed to accept NG sequence as PAM site but also combinations of NAC, NTG, NTT and NCG with lower nuclease activity (Nishimasu et al. 2018, Ren et al. 2019).

Newest alternative PAM site Cas9 variant was developed by Walton and colleagues in early 2020 and is called SpRY. What differentiates this variant from the earlier nucleases is the fact that SpRY nuclease is referred as near-PAMless, which Walton and colleagues explains to signify it can target nearly all possible PAM site nucleotide combinations. This is possible due to SpRY recognizing effectively only single guanine (NGN) as PAM site. SpRY also targets sites with NAN PAM site and even NCN and NTN combinations at lower effect affirming the fact SpRY nuclease is near-PAMless in function (Walton et al. 2020).

Cas9 proteins can be modified to have alternative functions other than original sequence targeting and cutting. Nuclease-deficient Cas9 (dCas9) is Cas9 protein that is mutated to lack endonuclease activity and thus does not cleave DNA (Qi et al. 2013). nCas9 is nickase variant that nicks only single strand at a time allowing more phased editing.

Base editing means changing single nucleotides in target sequence precisely without causing double-strand breaks and need for subsequent nonhomologous end-joining. Base editing was pioneered in 2016 by Komor and colleagues when they developed base editor that can change cytosine-guanine base pair (C·G) to adenine-thymine base pair (A·T). This was achieved by fusing Cas9 with cytidine deaminase enzyme that mediates conversion of cytosine to uracil. This anomalous uracil-guanine base pair (U·G) would normally lead to correction by uracil-DNA glycosylase that removes uracil from DNA but the base pair editor has attached additional uracil-DNA glycosylase inhibitor that cancels this effect leading to only option of correction by mismatch repair that allows desired outcome of A·T (Komor et al. 2016).

Base editor capable of reverse change from A·T to C·G was developed in 2017 by Gaudelli and colleagues by engineering synthetic transfer RNA adenosine deaminase that was fused with mutated Cas9 protein resulting in first adenine base editor (Gaudelli et al. 2017). This newfound ability to change single bases combined together with near-PAMless CRISPR-Cas9 system leads to intriguing possibility where almost all diseases resulting from point mutations could theoretically be cured in humans (Komor et al. 2016, Walton et al. 2020).

Prime editing is game changing new technology developed in 2019 by Anzalone and colleagues by fusing together dCas9 with engineered reverse transcriptase domain and prime editing guide RNA (pegRNA). pegRNA like all gRNAs leads the prime editor to desired sequence in genome but instead of double-strand break made by normal Cas9, the modified Cas9 has nickase activity (nCas9) and cuts only PAM site containing strand and produces single-strand cut. This cut end is recognized and bound to primer binding site in the end of pegRNA. This site then serves as a starting point for the following fused reverse transcriptase that copies new genetic information from the pegRNA into the target strand (Anzalone et al. 2019).

After transcription the editing site contains intermediate branched structure with overhanging 3'strand containing edited sequence and 5'strand containing unedited sequence. While the unedited 5'strand overhang is more likely to be hybridized to the complementary strand, 5'ends are also preferential substrate for specific endonucleases that remove them leading to higher chances of edit containing 3'end overhang to be hybridized to complementary strand. Heteroduplex caused by difference between original sequence and edited sequence is corrected by DNA repair machinery and this would naturally be equally likely to correct edited sequence to match original sequence as it is vice versa, but at this point prime editor also produces cut to the strand containing original sequence causing DNA repair machinery to favor uncut strand containing edited sequence as correcting template (Anzalone et al. 2019).

More complex binding specificity of prime editor compared to original CRISPR-Cas9 system leads to reduced amount of off-target mutations. Prime editors unparalleled versatility allowing it to perform all different types of DNA edits due to programmable pegRNA will likely elevate it as most prevalent editing technology for the future (Anzalone et al. 2019).

Cas proteins other than Cas9 offer novel new properties that can lead to even more diverse utilities when combined with techniques pioneered in Cas9. Remarkable addition to CRISPR toolbox is Cas12a (Cpf1) that produces double-stand breaks with staggered ends unlike blunt-ended cuts performed by Cas9. These staggered ends can greatly enhance chance of inserted gene attaching to target strand (Zetsche et al. 2015). Efficiency of Cpf1was enhanced in 2018 by Moon and colleagues by adding uridinylate-rich 3′-overhang to its crRNA. This addition, although not fully explained by the authors,

enhanced the molecules cutting activity by 13-fold and brought it close to Cas9 in efficiency (Moon et al. 2018).

2.6.2 CRISPR-Cas systems beyond gene editing

CRISPR-Cas systems are not limited to only cutting DNA or changing precise bases, they have several other functions that can be harnessed for human advantage. Fascinating function was found from CRISPR-Cas systems based on Cas13 and Cas14 proteins. CRISPR-Cas13a targets viral single-stranded RNA (ssRNA) and can be engineered to provide interference based immunity in plants against ssRNA viruses (Aman et al. 2018). CRISPR-Cas14a targets viral single-stranded DNA (ssDNA) and similarly can be engineered to provide interference immunity against ssDNA viruses in plants (Khan et al. 2019). These systems together could provide unprecedented possibilities in fight against common yield reducing plant viruses such as cassava brown streak virus (ssRNA) and cassava mosaic viruses (ssDNA).

Genomic imaging in living cells with fluorescent proteins is another area of research that CRISPR-Cas systems can improve. System called CRISPRainbow was developed in 2016 by Ma and colleagues that is based on dCas9 with engineered sgRNA able to attach fluorescent proteins to specific chromosomal loci thus enabling advanced simultaneous study of chromosome dynamics of multiple loci. Different combinations of two fluorescent proteins were attached to hairpin structures of sgRNA and with these all primary colors (RBG) were attained and when two colors were combined together the secondary colors (CMY) were also attained and presented in target cells. Plasmid expressing six different sgRNAs and thus enabling labelling of six targets simultaneously was developed to serve the CRISPRainbow system (Ma et al. 2016).

Epigenetic regulation is responsible for determining activity of all genes in an organism by different mechanisms such as DNA methylation and histone modifications that downregulate gene activity when generated into DNA strand and upregulate activity when removed. Because of this unfathomable extent to which epigenetic regulation determines biochemical activities in organism, its possibilities are yet to be discovered to full extent. Targeted DNA demethylation with CRISPR-Cas9 was first achieved in 2016 by Morita and colleagues by fusing inactive Cas9 nuclease (dCas9) with catalytic domain

from human originating demethylase TET1 (Morita et al. 2016). This was later modified to function in plants by Gallego-Bartolome and colleagues in 2018 when their team performed demethylation in *Arabidopsis* of the homeobox gene *FLOWERING WAGENINGEN* (*FWA*) that was activated as expected and resulted in late-flowering phenotype that was also heritable due to formation of a stable epiallele (Gallego-Bartolome et al. 2018).

3 RESEARCH OBJECTIVES

The research goal of this master's thesis was to discover efficient methods for CRISPR-Cas9 induced gene editing in the genus *Nicotiana* by comparing different methods and protocols developed and documented in previously published studies. Two approaches of gene editing were compared, CRISPR-Cas9 editing by *Agrobacterium* transformation and editing with CRISPR-Cas9 ribonucleoprotein complex. With each method different protocols were studied for optimal performance and verification of successful gene editing. Production of Cas9 protein was studied by using two different *E. coli* strains, vectors and protocols varying in production temperature and time. Verification of successful gene editing was studied with two different protocols for mutation detection, by using a commercial kit and by self-prepared celery extract containing natural single-strand specific nucleases.

4 MATERIALS AND METHODS

4.1 Strains

4.1.1 *Escherichia coli* strains

E. coli Rosetta(DE3) (Novagen) is a strain designed to express eukaryotic proteins from rare codons with enhanced efficiency. Transformed with plasmids pET28a-Cas9-His and pET28b-Cas9-His to produce strains Rosetta(pET28a-Cas9-His) and Rosetta(pET28b-Cas9-His), these are later referred as R(28a) and R(28b), respectively.

E. coli Lemo21(DE3) is a strain designed to express challenging proteins with adjustable expression levels controlled by L-rhamnose (Wagner et al. 2008). Transformed with plasmids pET28a-Cas9-His and pET28b-Cas9-His to produce strains Lemo21(pET28a-Cas9-His) and Lemo21(pET28b-Cas9-His), these are later referred as L(28a) and L(28b), respectively.

E. coli strain DH5α is multipurpose strain used for general cloning. It was used as host cell for commercial cloning vector pJET1.2.

4.1.2 *Agrobacterium* strains

C58C1(pGV2260), later referred to as GV2260, is an *Agrobacterium* strain carrying octopine type Ti-plasmid pGV2260. This plasmid has its transfer DNA region removed, which allows controlled agrotransformation with desired DNA insert (Deblaere et al. 1985). Strains GV2260(pDAR1) and GV2260(pDAR2), later referred as DAT1 and DAT2, were produced in collaboration with visiting exchange student Daria Pajak who transformed GV2260 strain with STU expression vector plasmids and selected them in her own research project and donated them to be used in agroinfiltration experiment for this thesis. The following sister strains of DAT1 and DAT2 were also used in the experiments: E1-2A (=DAT1), E1-2B, E1-3A, E1-3B and E2-2A (=DAT2), E2-2B, E2-3A, E2-3B.

Strains GV2260(pPOG1) and GV2260(pPOG2), later referred as POG1 and POG2, are improved versions of the previous with enhanced STU expression vector plasmids and were produced in collaboration with visiting exchange student Paloma Ortiz García who transformed and selected them in her own research project and donated them to be used in agrotransformation experiment for this thesis. See section 4.3.5 for description of pDAR1, pDAR2, pPOG1 and pPOG2.

Strain GV2260(pBin61-P19), later referred as P19, was used in *N. benthamiana* agroinfiltration as enhancer with DAT1 and DAT2. P19 is protein from tomato bushy virus (TBSV) that suppresses post-transcriptional silencing in plants caused by small interfering RNAs (siRNA). P19 forms complexes with siRNA and as a result effector complexes of the RNA-silencing machinery cannot work, which enables enhanced production of target protein by the primary vector (Voinnet et al. 2003).

4.1.3 *Nicotiana* strains

Nicotiana tabacum L. cv. 'Petit Havana SR1' (SR1) and *Nicotiana benthamiana* plants were obtained from constantly maintained and propagated university's laboratory stock. Strains have been used for several years and stock plants are sown from seeds originating from this pure line. *Nicotiana tabacum* is allotetraploid species containing genomes of *Nicotiana sylvestris* (Speg. & Comes) and *Nicotiana tomentosiformis* (Linné). *Nicotiana benthamiana* is allotetraploid containing genomes from two unidentified species from sections *Sylvestres* and *Noctiflorae* (Schiavinato et al. 2020). When sequence data was gathered and primers were designed, it was erroneously thought that *N. benthamiana* is a diploid species.

4.2 Target gene selection

4.2.1 *PDS1* and *PDS2*

PDS1 and *PDS2* are two loci encoding phytoene desaturase in genus *Nicotiana* that catalyzes lycopene formation in carotenoid synthesis. Mutant genotypes fail to produce carotenes, which leads to a photobleaching of chlorophyll allowing visual detection of mutated plants with distinctively white leaves and thus targeting these genes is practical

for study where effectiveness of different gene editing protocols is compared. *NtPDS* gene in *N. tabacum* was successfully targeted and mutated with CRISPR-Cas9 system in study by Gao and colleagues in 2015 where their team managed to produce albino plants as result of loss-of-function mutation in this gene due to frameshift mutation (Gao et al. 2015).

4.2.2 Sequence data

Genomic sequences for PDS were searched using The Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) (National Center for Biotechnology Information, U.S. National Library of Medicine). *N. benthamiana* PDS sequence was found from genomic scaffold Niben101Scf01283 (Sol Genomics Network) by using query sequence DQ469932.1 (benthamiana PDS mRNA). *N. tabacum* PDS sequences were analyzed as combination from *N. sylvestris* genomic scaffold NW_009521166 (NCBI) by using query sequence XM_009805088.1 (sylvestris PDS mRNA) and *N. tomentosiformis* genomic scaffold NW_008898264 (NCBI) by using query sequence XM_009615718.2 (tomentosiformis PDS mRNA).

4.3 Primer design and plasmids

4.3.1 Primer design

Primers designed to amplify the Cas9 edited target sites had two main uses in the experiments. First, verification of gene editing in agroinfiltration was done from PCR products produced from transformed plants and were subjected to mutation detection and comparison to PCR products from unaltered plants. Second, target sites were cloned into commercial cloning vector pJET1.2 for sequencing.

Phytoene desaturase is translated from two loci *PDS1* and *PDS2* thus both loci need to be targeted with individual primer set of forward and reverse primers. *PDS1* and *PDS2* sequences of *N. tabacum* (*sylvestris* and *tomentosiformis* genomes) and *N. benthamiana* were aligned to find corresponding sequences from which compatible primer sequences and Cas9 target sequences were possible to design.

Primers for *N. benthamiana*

At this point it was believed that *N. benthamiana* is diploid species, thus only single copies of the PDS loci were retrieved from sequence data.

Benthamiana *PDS1*		
Forward	GER1134	GGAACTGAAAGTCAAGATGGTCG
Reverse	GER1135	CACCGATTTGGGACAATTCAACAGT
PCR fragment	622 bp	

Benthamiana *PDS2*		
Forward	GER1117	GCTGTAAATCATGTTAGGGACCTG
Reverse	GER1118	CATCCATAAGTTCCTCATTCAAGGG
PCR fragment	657 bp	

Primers for tetraploid species *N. tabacum*

Sylvestris *PDS1*		
Forward	GER1130	GGGAACTGAAAGTCAAGATGGTCA
Reverse	GER1131	GTGGGACAACTCAACAGTATAAGAG
PCR fragment	632 bp	

Sylvestris *PDS2*		
Forward	GER1136	CTGACATATTGGTGCAGGAAACTTATTTC
Reverse	GER1137	AACACAATACCGGAGGACCACATAAA
PCR fragment	611 bp	

Tomentosiformis *PDS1*		
Forward	GER1132	GGAACTGAAAGTCAAGATGGTCG
Reverse	GER1133	TAGATATGTATATATAGATATGTATATAA-
PCR fragment	623 bp	ATGGAGTGACG

Tomentosiformis *PDS2*		
Forward	GER1138	TGACATATTGGTGCAGGAAACTTATGAG
Reverse	GER1139	GAAAGTAACACAAAAACAGAGGACTACA-
PCR fragment	611 bp	TGTT

N. benthamiana PDS2 had different original primer design that was combination from *N. benthamiana* specific forward primer GER1140 (ACATATCGGTGCAGG-AAACTTATGAG) and reverse primer GER1139 that was designed as common reverse primer for both *N. benthamiana* and *N. tomentosiformis PDS2*. Combination was not functional for *N. benthamiana* thus new primer pair GER1117 + GER1118 was designed. Running number of the new primer pair is less than others due to accounting reasons.

4.3.2 PCR program

PCR program was set with following steps.

1. Initial denaturation at 95 °C, 3 minutes
2. Denaturation at 94 °C, 1 minute
3. Annealing at 55 °C, 30 seconds
4. Extension at 72 °C, 1 minute
5. Multiplying with “go to step 2”, 29 times
6. Cooling at 15 °C, 18 hours

4.3.3 Cas9 target HinfI sites

HinfI cutting and recognition site GANTC in the target gene sequence was chosen as Cas9 target and successful mutation in this sequence could be verified by ineffective HinfI cutting in addition to T7E1 mutation detection and white shoots in stable transformants. Cas9 needs protospacer adjacent motif (PAM site) following targeted gene sequence for recognition and functioning. PAM site is marked by two consecutive guanines following any nucleotide thus HinfI sites followed by 5'-NGG-3' were chosen.

PDS1

N. sylvestris TTGGTAGTAGCGACTCCATGGGG

N. tomentosiformis TTGGTAGTAGCGATTCCATGGGG

N. benthamiana TTGGTAGTAGCGACTCCATGGGG

PDS2

N. sylvestris GAACATATTGAGTCAAAAGGTGG

N. tomentosiformis GAACATATTGAGTCAAAAGGTGG

N. benthamiana GAACATATTGAGTCAAAAGGTGG

4.3.4 Plasmids pET28a-Cas9-His and pET28b-Cas9

Plasmids designed for Cas9 production are constructed on Novagen pET-system plasmid vectors pET-28a and pET-28b. Both vectors contain N- and C-terminal His-tag that enables protein identification and purification of the Cas9 end-product based on binding to Ni^{2+}. Two different plasmids with inserted Cas9 sequence were used, these are pET28a-Cas9-His (Liang et al. 2017) and pET28b-Cas9-His (Gagnon et al. 2014) that have different versions of His-tagged Cas9 (Figure 14, appendices). pET28a-Cas9-His version has normal Cas9 with two SV40 nuclear localization signals (NLS) allowing nuclear import of Cas9 protein. Addition of multiple SV40 NLS signals, instead of a single NLS operating in manner of individual switch, can allow more gradual regulation of nuclear protein levels through import level alteration (Luo et al. 2004). pET28b-Cas9-His has only one SV40 NLS signal and human codon-optimized Cas9 (hCas9).

4.3.5 Plasmids pTX171, pDAR1, pDAR2, pPOG1 and pPOG2

STU expression vector produces both Cas9 protein and sgRNA in single gene expression event. Plasmid vector pTX171 is ready-made STU expression vector for CRISPR-Cas9 that contains Cas9 sequence and empty scaffold section for designated single-guide RNA. When equipped with appropriate sgRNA and transformed into *Agrobacterium* it performs gene editing in recipient cell (Tang et al. 2016).

Visiting exchange student Daria Pajak produced plasmids containing sgRNA for *Nicotiana PDS1* (pDAR1) and *Nicotiana PDS2* (pDAR2) from pTX171vector with

protocol provided by Tang and colleagues. These plasmids convey hygromycin resistance to recipient cells as selection marker. Cas9 target sequences added in the sgRNA are HinfI cutting sites in *PDS1* and *PDS2*. These sites are identical between *N. benthamiana* and *N. tabacum* thus editing should work in both species (Figure 15, appendices).

Plasmids pDAR1 and pDAR2 were later modified by visiting exchange student Paloma Ortiz García to contain kanamycin resistance as selection marker instead of hygromycin. These modified plasmids with kanamycin selection marker are pPOG1 and pPOG2 (Figure 16, appendices).

4.3.6 Plasmid pJET1.2

pJET1.2 is a commercial cloning vector included in CloneJET PCR Cloning Kit (ThermoFisher). It was used to study *PDS* genes of university stock plants by cloning primer produced sequences into this plasmid. Plasmids were then sent to sequencing.

4.3.7 Plasmids pHTT837 and pHTT838

pHTT837 and pHTT838 are plasmids that were used to produce PCR products later used as test material for mutation detection with celery extract. Both plasmids contain the same insert, where one restriction site in each has been eliminated by filling-in. When PCR products of their inserts are combined, they form heteroduplexes with two mismatch bubbles that CEL I cuts and this can be detected in gel electrophoresis as positive control for mutation (Figure 17, appendices).

4.4 Culture Media and Conditions

4.4.1 Bacterial culture media

All *E. coli* and *Agrobacterium* media were prepared from common premade agar solidified L-broth medium that was sterilized and melted in autoclave and cooled before adding strain specific antibiotics. Standard 30 ml plates were poured in laminar flow cabinet and stored in cold room for later use. *E. coli* plates were cultured in +37 °C growth chamber overnight. *Agrobacterium* plates were cultured in +28 °C darkened growth

chamber for three days. Liquid cultivation of *E. coli* and *Agrobacterium* was done with common premade liquid L-broth medium in standard 5 ml glass tubes with or without added antibiotics and overnight shaking for *E. coli* and *Agrobacterium*.

Selection antibiotics used for *E. coli* were kanamycin (50 µg/ml) and chloramphenicol (30 µg/ml). Selection antibiotics used for *Agrobacterium* were rifampicin (100 µg/ml), carbenicillin (100 µg/ml) and kanamycin (100 µg/ml).

4.4.2 *Nicotiana tabacum* culture medium

Transformed *N. tabacum* leaf pieces were cultivated in standard MS-medium that was prepared with 4.4 g Murashige & Skoog salt mixture + vitamins, 1.0 mg/L BAP, 20 000 mg/L sucrose and filled to 1000 ml with RO water and adjusted to pH 5.7 with potassium hydroxide. 100 ml was set aside as co-cultivation liquid for agrotransformation before 7.2 g of commercial bacto agar was added to remaining 900 ml before autoclaving 20 minutes at 120 °C. After cooling, control 0-plates containing no antibiotics were poured in necessary quantity before first adding claforan (500 mg/l) to remaining media and control plates without selection were poured. Finally, the selection antibiotic was added to remaining liquid and remaining plates were poured. In the first experiment selection antibiotic was hygromycin (50 mg/l) and later kanamycin (100 mg/l).

4.5 Transformation

4.5.1 *E. coli* transformation

Heat shock protocol was executed by thawing Eppendorf tube containing competent cells on ice for 20 minutes, after which 5 µl of plasmid DNA was added to cell suspension and left on ice for 30 minutes. Heat shock was given at +42 °C for 45 seconds and tube cooled on ice for 2 minutes. 1 ml of L-broth was added to tubes and cells were given 60 minutes recovering time at +37 °C in warming cabinet before spinning 3 minutes at 2500 xg. Supernatant was poured away leaving approximately 200 µl of liquid to which cells were resuspended and plated to selection plates containing kanamycin (50 µg/ml) and chloramphenicol (30 µg/ml).

Electroporation protocol was executed by thawing Eppendorf tube containing competent cells on ice for 20 minutes, after which 5 µl of plasmid DNA was added to cell suspension and left on ice for 1 minute. Using chilled tips (+4 °C) cell suspension was moved to chilled cuvettes and electroporation was performed with program 200 ohms, 25 µF, 2.5 kV. Cells were rinsed to 12 ml Falcon tubes with 500 µl of L-Broth and incubated 60 minutes at +37 °C with shaking before spinning 3 minutes at 2500 xg. Supernatant was poured away leaving approximately 200 µl of liquid to which cells were resuspended and plated to selection plates containing kanamycin (50 µg/ml) and chloramphenicol (30 µg/ml).

The heat shock protocol was used for both Rosetta and Lemo21 strains while electroporation protocol was tested only for Rosetta strain. Four colonies were selected from each unique transformation for verification done by PCR.

Plasmids were isolated with GenElute Plasmid MiniPrep Kit (Sigma-Aldrich) according to protocol from overnight 5 ml cultures. Protocol's optional wash was performed with 500 µl of Optional Wash Solution. Resulting plasmid DNA was eluded to 80 µl of TE-buffer.

Plasmids were digested with PstI and PvuII for 15 minutes at 37 °C in solution of 2 µl 10X buffer, 15 µl MQ water, 3 µl DNA, 0.3 µl PstI and 0.3 µl PvuII before running in 1.5% gel.

Second transformation experiment of *E. coli* was done with primer produced *PDS1* and *PDS2* fragments processed to be blunt-ended with CloneJET PCR Cloning Kit to produce plasmids suitable for sequencing company. Cloning was done according to protocol on commercial pJET1.2 plasmid included in the kit.

4.6 CAS9 production and extraction

4.6.1 Cas9 production in *E. coli*

Cas9 production protocol was based on pET System Manual 11th Edition (Novagen) and similar in principle in all production experiments differing only in which *E. coli* strains were used, induction temperature (+16 °C, +22 °C, +28 °C or +37 °C) and time (2 h, 3 h, 4 h or overnight) and lysing method (lysozyme, B-PER or CelLytic). Below is described basic method which was modified depending on experiment.

Experiment started from fresh overnight plates of *E. coli* strains R(28a), R(28b), L(28a) and L(28b) from which 2 ml L-broth (kanamycin (50 µg/ml), chloramphenicol (30 µg/ml) precultures were prepared in morning. Precultures were grown at +37 °C with shaking until absorbance OD600 = 1.0 was reached. 0.5 ml of preculture was used to inoculate warm 3 ml L-cam-kana cultures. Four of these 3.5 ml cultures were prepared per strain and grown at +37 °C with shaking until absorbance OD600 = 1.0 was reached.

From these four cultures per strain, two were moved to chosen induction temperature of +22 °C with one tube as control and other tube with IPTG added to 0.5 mM and other two kept at +37 °C with one tube as control and other tube with added IPTG (0.5 mM). Cultures were grown with shaking for 2 hours (+37 °C) and 4 hours (+22 °C). Absorbances at OD600 were measured before cells were collected.

1 ml of cells were collected from each control culture into 2 ml Eppendorf tubes for total protein analysis. Two Eppendorf tubes were collected from each IPTG containing culture with 1 ml of cells as from one tube total proteins were analyzed and from second tube soluble and insoluble proteins were later measured separately. Cells were spun at 10 000 rpm for 5 minutes and supernatant poured away before second short 10 seconds spinning, after which rest of supernatant was pipetted away and samples frozen for next day.

4.6.2 Protein gel analysis

Frozen sample pellets from Cas9 production protocol were resuspended in 60 µl of plant extraction buffer (50 mM Tris-Cl, pH 7.5 with Roche protease inhibitor cocktail tablets

added by protocol). Total protein samples were kept on ice while 6 µl of lysozyme (10 mg/ml) was added to samples with protein fractioning. Samples with lysozyme were incubated at room temperature until seemingly viscous, after which they were exposed three times to cycle of freezing on dry ice and thawing and vortexing. Samples were spun at full speed for 10 minutes, after which supernatant containing soluble proteins was moved to new tube and pellet containing insoluble proteins was resuspended to 66 µl of plant extraction buffer. At this step from each of four *E. coli* strains there are eight samples, four from each different induction temperature: total protein control without IPTG, total protein from IPTG induced culture, insoluble protein from IPTG induced culture and soluble protein from IPTG induced culture.

20 µl of 4X SB was added to each tube and incubated 5 minutes at 95 °C, after which tubes were spun 5 minutes at full speed. Loading amounts for protein gel analysis were calculated as 40 µl for OD600=1.0 and for actual loading amount 40 µl was by each sample's true OD600 reading. Maximum of 20 µl could be loaded into gel thus loadings were downgraded by half. Mini-PROTEAN TGX Stain-Free Precast Gels (Bio-Rad Laboratories) were used in experiments. Protein gels were run with 5 µl of protein gel size controls. After running gels for one hour they were rinsed three times in shaking for 10 minutes with RO-water and stained overnight with Coomassie Brilliant Blue.

4.6.3 Protein purification

His SpinTrap (GE Healthcare) single-use spin columns were used to purify histidine-tagged Cas9 protein. Purification principle is based on metal ion affinity chromatography that binds proteins with polyhistidine motif into column while other non-histidine-tagged proteins flow through. Target protein is then collected with suitable elution buffer. Buffers needed in purification were prepared by protocol.

Binding buffer was prepared following kit instructions containing 20 mM sodium phosphate pH 7.5, 500 mM NaCl, 20 mM imidazole and prepared in MQ-water.

Elution buffer was prepared following kit instructions containing 20 mM sodium phosphate pH 7.5, 500 mM NaCl, 500 mM imidazole and prepared in MQ-water.

Storage solution was removed from columns shaking columns until medium was resuspended, after which cap was loosened by one-quarter turn and bottom closure removed. Columns were spun in microcentrifuge at 100 xg for 30 seconds to remove storage buffer. Cap was removed, and columns were equilibrated with 600 µl of binding buffer by spinning again for 30 seconds. Samples were lysed and cleared by centrifugation. 600 µl of sample was added to equilibrated column at a time and spun at 100 xg for 30 seconds. This was repeated if entire sample did not fit the column in one time until entire sample had been poured to column, after which column was washed with 600 µl of binding buffer. Samples were eluded twice with 200 µl of elution buffer by spinning columns at 100 xg for 30 seconds and collecting purified sample. From the two 200 µl eluted samples first one theoretically contains majority of Cas9, but these samples were combined to one 400 µl sample.

PD MiniTrap G-25 (GE Healthcare) single-use gravity columns were used to remove imidazole and salt compounds from final protein samples and to change His SpinTrap elution buffer to Cas9 storage buffer that was prepared beforehand.

Cas9 storage buffer contained 20 mM Hepes-buffer pH 7.5, 150 mM KCl, 1 mM DTT prepared in MQ-water with 50% glycerol added only before storage.

Columns were assembled and prepared by removing cap and pouring away storage solution. Bottom cap was removed, and column equilibrated with Cas9 buffer without glycerol by filling column and letting it flow to bottom portion of column, this was repeated twice. Flow through was discarded and sample (400 µl) was added to column. After the sample had flown to the bottom portion of column, 100 µl of Cas9 buffer was added and again waited until immersed, after which flow through was discarded. This was done to even out the amount of liquid entering the column bottom to 0.5 ml, which is the actual volume of liquid column bed holds at a time. Following 400 µl of flow through will contain purified protein sample. Collection tube was placed under the column and sample eluted with 1.0 ml of Cas9 storage buffer. 250 µl was collected at a time and tube changed until no more drops formed. Protein concentrations were measured from each tube and majority of sample protein located to first two 250 µl fractions that were combined to a single sample. Purified protein was studied on protein gel with size marker and glycerin added to remaining sample.

4.6.4 Verification of product by western blotting and antibody incubation

Western blotting was done for samples in Bio-Rad ready-made protein gel by rinsing individual gels 10 minutes in transfer buffer, after which proteins were transferred to Amersham Hybond-ECL nitrocellulose membrane for 30 minutes using Trans-Blot Turbo Transfer System (Bio-rad). Membranes were stained with Ponceau S for 1 hour, after which they were rinsed with tap water and photographed. Membranes were rinsed in Tween-Tris-buffered saline (tTBS) (20 mM Tris-Cl pH 7.5, 150 mM NaCl and 0.1% Tween 20 made in RO water) for 5 minutes, after which they were shaken in 25 ml of milk-based blocking buffer (5% nonfat dry milk powder in tTBS) for 1 hour. Membranes were rinsed in tTBS for 10 minutes. Primary antibody serum His-Tag antibody #2365 (Cell Signaling Technology) was mixed in 10 ml of primary antibody dilution buffer (5% Bovine Serum Albumin in tTBS). Membranes were rolled protein side facing center into 50 ml Falcon tubes and antibody was added to tubes. Tubes were mixed overnight in roller at +4 °C. Membranes were rinsed 3 times with tTBS for 10 minutes. Secondary antibody serum Anti-rabbit igG, HRP-linked antibody #7074 (Cell Signaling Technology) was added in 10 ml blocking buffer and distributed to membranes, after which they were shaken 1 hour at room temperature. Membranes were rinsed 4 times with tTBS for 10 minutes. Excess water was removed from membranes and they were moved to glass plate, after which 1 ml of Amersham ECL Prime Western Blotting Detection Reagent was spread to membranes using plastic roller. Excess liquid was removed, and plastic film spread over membranes. Membrane plate was placed into radiation cassette and exposed and developed to x-ray film in dark room. Exposure times of 1 second, 2 minutes, 30 minutes and overnight were used and developed into individual x-ray films.

4.7 Tobacco *PDS1* and *PDS2* cloning and sequencing

4.7.1 Cloning

University's stock plants were studied whether they have single nucleotide polymorphisms in their *PDS* genes compared to published sequences. *PDS* gene fragments were produced with corresponding primers and these PCR products were cloned to pJET1.2 vector according to kits protocol.

Tetraploid species *N. benthamiana*

N. benthamiana	locus *PDS1*	primers GER1134+ GER1135
N. benthamiana	locus *PDS2*	primers GER1117+ GER1118

Tetraploid species *N. tabacum*

N. sylvestris	locus *PDS1*	primers GER1130+ GER1131
N. sylvestris	locus *PDS2*	primers GER1136+ GER1137
N. tomentosiformis	locus *PDS1*	primers GER1132+ GER1133
N. tomentosiformis	locus *PDS2*	primers GER1138+ GER1139

4.7.2 Sequencing

Plasmids were isolated from *E. coli* strains transformed with all different *PDS* gene variants from different tobacco species by using GeneJET Plasmid Miniprep Kit (ThermoFisher) and send to Macrogen for sequencing following their sample preparation protocol.

4.8 CRISPR-Cas9 experiments

4.8.1 Agroinfiltration on *Nicotiana benthamiana*

Agroinfiltration experiment started from *Agrobacterium* overnight cultures of 5 ml L-broth with shaking at +28 °C. Cultures were transferred to 15 ml Falcon tubes and spun 10 minutes 4000 rpm, after which supernatant was removed and tubes were spun again for 10 seconds. Remaining supernatant was removed with pipette and pellet was resuspended in 2 ml of Mg-MES (200 µM acetoryringone, 10 mM $MgCl_2$, 10 mM MES-buffer, pH 6.0). 200 µl of *Agrobacterium* suspension was diluted to 3 ml of Mg-MES and absorbance measured at 600 nm (O.D. 600). Depending on absorbance reading dilution was adjusted either with undiluted bacterium solution or Mg-MES until reading of O.D. 600 = 0.500 was achieved. Bacterium suspension was left in room temperature for three hours before agroinfiltration that was done with plastic syringe on two topmost leaves of ten-leaf stage *N. benthamiana* plants.

DNA was extracted from leaf samples using Miniprep II protocol (Dellaporta et al. 1983). Around 100 mg of leaf tissue (exact weight recorded) was collected on round-bottom Eppendorf tube and two steel beads were added. Tubes were frozen on liquid nitrogen and milled to powder using ball mill homogenizer. Leaf powder was resuspended into 600 µl of extraction buffer (100 mM Tris-Cl, 50 mM EDTA, 500 mM NaCl, 20 mM 2-ME, 67 µg/ml RNase, pH 8.0). Samples were moved to 1.5 ml Eppendorf tubes and 40 µl of 20% SDS was added and mixed well. Tubes were incubated at +65 °C for 10 minutes and occasionally mixed. 200 µl of 5M potassium acetate was added to lysed samples and mixed well. Tubes were spun at maximum speed for 10 minutes at room temperature. Supernatant was collected, and 0.7 volumes of cold isopropanol was added. Tubes were spun at maximum speed for 10 minutes at +4 °C. Supernatant was discarded, and pellet washed with 0.5 ml of cold 70% ethanol by spinning at maximum speed for 10 minutes at +4 °C. Supernatant was discarded and drops spun quickly to bottom and removed with pipette. Pellet was not allowed to dry before resuspending overnight in 300 µl TE (10 mM Tris-Cl, 1 mM EDTA, pH 8.0).

4.8.2 Agrotransformation on *Nicotiana tabacum*

Agrobacterium overnight cultures of 5 ml L-broth without antibiotics from single colonies were established on previous day. Medium sized young leaf of *N. tabacum* 'SR1' was selected and sterilized using wash protocol of one minute in 70% ethanol, rinsing in sterile water, ten minutes in 1:4 diluted Na-hypochlorite and final rinsing in sterile water for five times. Bleached edges were removed and 10 small pieces (0.5 cm x 0.5 cm) were cut per experiment part. Control plates without *Agrobacterium* (0-control, claforan control and selection control) were filled immediately with leaf pieces. Second batch of leaf pieces was cut for *Agrobacterium* infection and submerged in overnight cultures diluted with 30 ml of co-cultivation MS liquid poured on petri dish. After 5 minutes leaf pieces were dried on sterile Whatman paper and moved onto 0-plates and incubated in dark cabinet at room temperature for two days to cause *Agrobacterium* infection to occur. Next infected leaf pieces were moved to selection plates containing claforan (500 mg/l) and hygromycin (50 mg/l) or kanamycin (100 mg/l) and kept in standard growing room conditions and changed to fresh plates every second week while observing shoot growth.

4.9 Mutation detection

4.9.1 T7 endonuclease 1 assay

Mutation detections were done using EnGen Mutation Detection Kit (NEB) according to protocol. Mutation detection was done for PCR products obtained from agroinfiltration experiments' samples that were first denatured and annealed to form heteroduplexes between non-mutated and mutated PCR products. This is important step as this mismatch between DNA strands forms "bubble" that functions as cutting site for T7E1. Annealing was performed in microcentrifuge tubes with 19 µl reaction containing 5 µl PCR product (~250 ng DNA), 2 µl 10X NEBuffer2 and 12 µl MQ-water. Samples were heat treated in PCR machine by heating them to 95 °C for 10 minutes, after which they were slowly cooled to room temperature.

Heteroduplex digestion was done by adding 1 µl of EnGen T7 Endonuclease I to annealed 19 µl samples that were mixed well and spun down. Samples were incubated at 37 °C for 15 minutes, after which 1 µl of Proteinase K was added and samples incubated at 37 °C for 5 minutes to terminate T7E1. Samples were analyzed in gel electrophoresis using 1.5% agarose gel.

4.9.2 Celery extract

Celery (*Apium graveolens subsp. dulce* (Mill.) Schübl. & G. Martens) synthesizes endonuclease CEL I that acts as natural mismatch cleavage enzyme in same manner as T7E1. Crude extract of CEL I can be made directly from celery plants with basic laboratory equipment and by using celery juice extract protocol (Till et al. 2003, Till et al. 2004). Using self-prepared crude extract makes mutation detection endonucleases available with very low cost. Crude extract was prepared from 400 g of commercial celery stalks and amount of achieved extract enables up to 10 000 mutation detection assays. Cost per single mutation detection using celery extract is insignificant compared to commercial T7E1.

Extraction was done in cold room and started from celery preparation that was done by separating and rinsing celery stalks and removing leaves and bottom white parts of

petioles. Green pieces of stalks were liquified using standard kitchen juicer and total of 200 ml was achieved. Celery juice was added to beaker with stirrer and Tris-buffer with PMFS was added to achieve concentration of 0.1 M Tris HCL, pH 7.7, 100 µM PMSF. The juice was transferred to a plastic bottle suitable for centrifugation and the bottle was spun at 2600 xg for 20 minutes at 4 °C. Supernatant was carefully poured to new beaker and pellet was discarded. In a cold room using stirrer, 27.36 g (144 g/l) ammonium sulfate was added to liquid (190 ml), stirred for 30 minutes and spun at 15 300 xg for 40 minutes at 4 °C. Supernatant was carefully poured to a beaker and the pellet was discarded. In a cold room using stirrer, 74.18 g (390 g/l) ammonium sulfate was added to the liquid (190 ml,) stirred for 30 minutes and spun at 15 300 xg for 90 minutes at 4 °C. Now the supernatant was discarded, and the pellet resuspended in 6 ml of Tris-buffer (0.1 M Tris HCL, pH 7.7, 100 µM PMSF). Original suggestion in protocol is to resuspend pellet into one tenth of original liquid volume but a more concentrated product was pursued. Liquid was dialyzed in Spectra/Por 10 000 MWCO tube against Tris-buffer (0.1 M Tris HCL, pH 7.7, 100 µM PMSF) in cold room. 4 liters of Tris-buffer was used each time and using hour interval, buffer was changed seven times. Final product (crude celery extract) was transferred to 15 ml tube and storage buffer additives were added (6 µl Triton X-100, 60 µl 1M KCl, 60 µl 1M $MgCl_2$, 60 µl 100x BSA stock).

5 RESULTS AND DISCUSSION

5.1 Primer design and functionality

Primers were designed for two *PDS* gene variants from three different genomes thus they were studied extensively to ensure appropriate functionality. Both *N. tabacum* and *N. benthamiana* are allotetraploid species containing two genomes from ancestral species but *N. benthamiana* was misidentified as diploid species when primers were designed, which lead to addressing only three genomes in these experiments.

Expected size of PCR fragments from different tobacco *PDS* genes are 611-657 bp in size (see 4.3.1).

Original primers (Figure 3A) functioned well providing PCR fragment of expected size for both *PDS1* and *PDS2* loci of *N. tomentosiformis* and *N. sylvestris* in N. *tabacum.* Primers for *N. benthamiana PDS1* functioned similarly well but *PDS2* primers did not yield PCR product.

New primer pair (Figure 3B) was designed for *N. benthamiana PDS2* and tested in combination with old primers. New primers functioned well providing PCR fragment of expected size and yielded PCR product also when old forward primer was paired with new reverse primer.

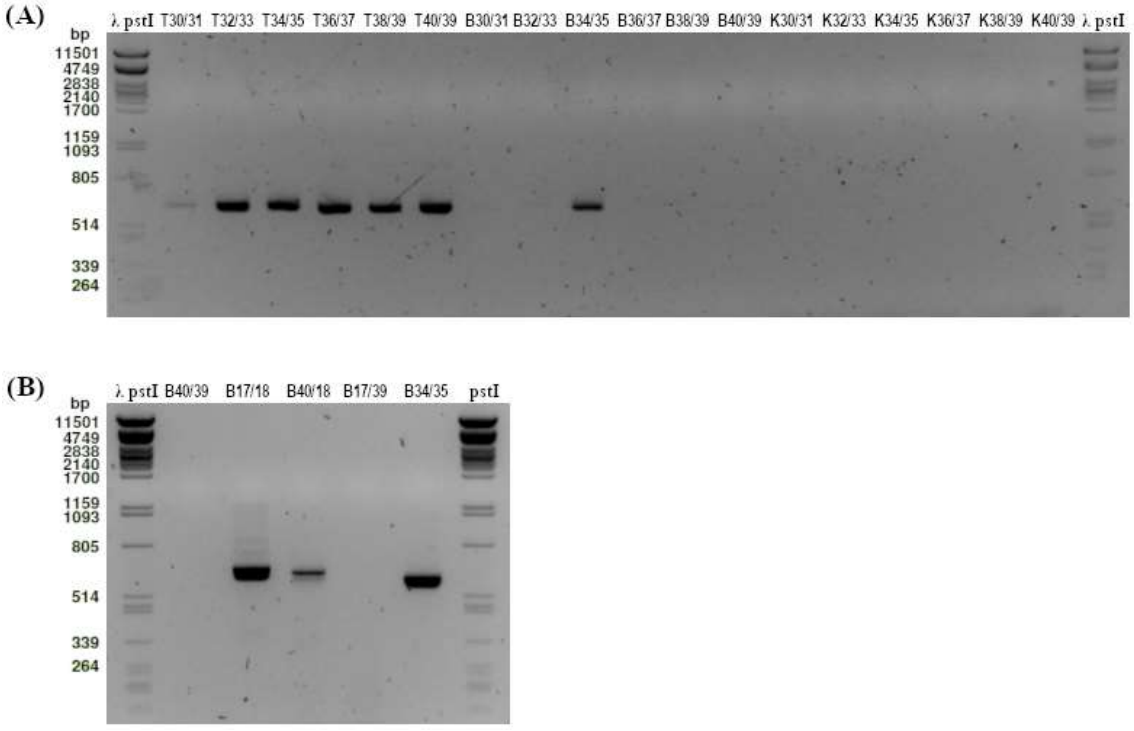


Figure 3. PCR products made from genomic DNA of *N. tabacum* and *N. benthamiana* with primers designed to target *PDS1* and *PDS2* loci. Sample names are composed from T = tabacum, B = benthamiana, K = water control and numbers indicating the primer pair used in PCR (last two digits of primer numbers). **(A)** Products from first PCR reaction using original primer set. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. **(B)** Products from second PCR reaction using new *N. benthamiana* primers in combination with original primers. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. Expected sizes of bands are ca. 600 bp (see text).

5.1.1 Difference in efficiency

Primers for *NbPDS1*, *NsPDS1*, *NsPDS2*, *NtPDS1* and *NtPDS2* functioned well and PCR products were obtained normally. *NsPDS1* PCR product showed lower concentration in gel compared to PCR products made with other primers. Sequence data (Figure 20, appendices) from PCR products made with *NsPDS1* primers GER1130 and GER1131 reveals that polymerase enzyme has copied two additional base pairs (TT) before forward primer sequence indicating inaccurate primer binding that could have led to problems in PCR reaction thus leading to lower concentration of final PCR product.

Original primers for *NbPDS2* (GER1140 + GER1139) were defective, which is likely due to difference in sequence between university's *N. benthamiana* plants and plants used for published genomic sequence of *N. benthamiana* from which primers were designed.

Sequence data (Figure 19, appendices) from successful PCR products made with new GER1117 + GER1118 primer pair reveals consecutive two base pair difference at the end of binding sequence for GER1140 in university's *N. benthamiana* plants, which has likely prevented polymerase binding. *NbPDS1* similarly (Figure 18, appendices) has two base pair difference at the end of binding sequence for GER1134 but not in consecutive order thus likely causing less disruption for polymerase binding.

5.2 *PDS1* and *PDS2* cloning and sequencing

PDS1 and *PDS2* gene fragments produced from *N. benthamiana* and *N. tabacum* with their corresponding primers were inserted to pJET1.2 plasmids, which were then transformed into *E. coli* strain DH5α. With these strains the gene fragment containing plasmids were multiplied and later extracted, after which the plasmids were sent to sequencing to compare genetic background of university's stock plants against publicly available sequences from *PDS1* and *PDS2* genes.

5.2.1 pJET1.2 transformation

In the beginning of this experiment transformation was problematic and failed twice with standard heat shock and electroporation with dialysis methods. It was hypothesized to result from impure PCR products and thus High Pure PCR Product Purification Kit was used before blunting reaction to ensure contaminants would not spoil transformation efficiency.

Digesting non-transformed pure pJET1.2 plasmid (2974 bp) with PstI and PvuII results 2057 bp, 522 bp and 395 bp fragments. 395 bp fragment contains cloning site for blunt-end PCR product thus size of gel fragment indicating successful transformation is 395 bp + size of PCR product. PCR fragments from different tobacco *PDS* genes 611-657 bp in size (see 4.3.1) thus size of combined fragment is around 1000 bp in size (1006-1052 bp).

Digesting non-transformed pure pJET1.2 plasmid with PvuI and HindIII results 1741 bp and 1223 bp fragments. 1223 bp fragment contains cloning site for blunt-end PCR product thus size of gel fragment indicating successful transformation is 1223 bp + size of PCR

product. Combined fragment including PCR products from this digestion is around 1800 bp in size (1834-1880 bp).

First transformation attempt using heat shock method with PCR samples purified with High Pure PCR Product Purification Kit before blunting reaction gave good results with uniform digestion bands from strains transformed with fragments from benthamiana *PDS1*, tomentosiformis *PDS1* and *PDS2* and sylvestris *PDS2*. Samples from benthamiana *PDS2* and sylvestris *PDS1* gave solely incorrect results with two large bands at ~2000 and ~5000 bp or smear-like band spreading between 2000 to <5000. Repeated gel run with sharper image (Figure 8A) was made from uniform strains.

Second transformation attempt using heat shock method with purified PCR samples gave good results with uniform digestion bands from strains transformed with fragments from sylvestris *PDS1*. Strains transformed with fragments from benthamiana *PDS2* gave several different versions of digestion bands including previous digestions with two large bands at ~2000 and ~5000 bp and smear-like band spreading between 2000 to <5000. Repeated gel run with sharper image (Figure 8B) was made from strains selected for later inspection including assumed incorrect digestions B2/4 and B2/5.

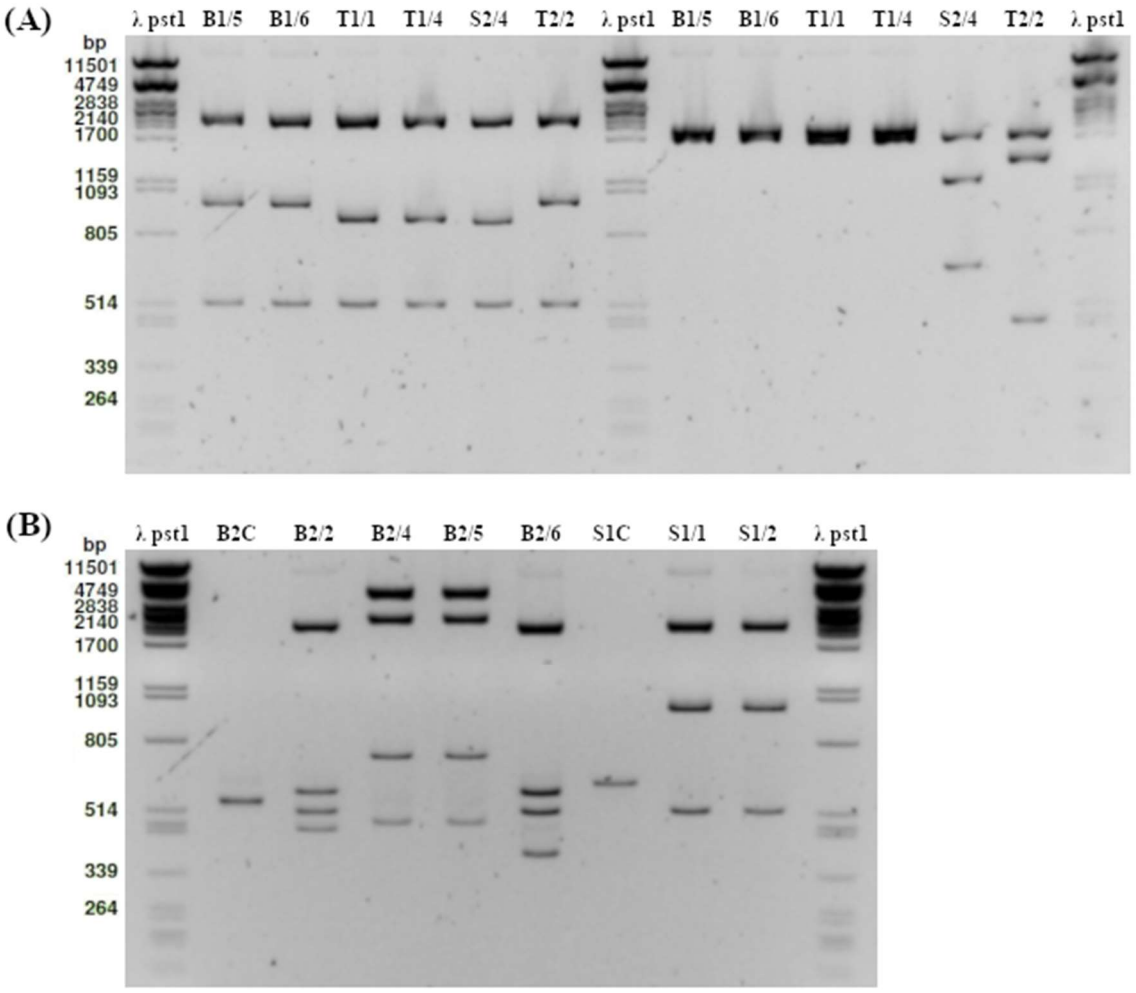


Figure 8. Digested plasmids from *E. coli* strain DH5α transformed with pJET1.2 plasmids containing PCR fragments form different versions of *PDS1* and *PDS2*. Sample names are composed from B = benthamiana, T = tomentosiformis, S = sylvestris, 1 = *PDS1* and 2 = *PDS2* and /1 - /6 responding to selection plate sector. B2C = benthamiana *PDS2* fragment and S1C = sylvestris *PDS1* fragment as controls in second gel image. **(A)** Digestion products of strains selected from first transformation attempt. Samples on left side of image were digested with PstI/PvuII and samples on right side with PvuI/HindIII. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. **(B)** Digestion products of strains selected from second transformation attempt. Samples were digested with PstI/PvuII. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. See text for expected fragments.

5.2.2 Successfully transformed pJET1.2 strains

Predicted PstI/PvuII digested fragment around 1000 bp in size indicating successful transformation is seen in samples B1/5 and B1/6 (N.b. *PDS1*), S1/1 and S1/2 (N.s. *PDS1*), S2/4 (N.s. *PDS2*), T1/1 and T1/4 (N.t. *PDS1*) and T2/2 (N.t. *PDS2*).

B2/2 and B2/6 gave different sized fragments compared to other PstI/PvuII digested samples but have correct insert in them. This is due to PstI cutting site in *N. benthamiana PDS2* fragment. This cutting site results combined fragment (1006-1052 bp) splitting into two fragments of 629 bp and 423 bp in size. These fragments are seen in B2/2 but in B2/6 second fragment is smaller (354 bp) due to deletion in PCR product.

In the first experiment (Figure 8A) samples were also digested with PvuI and HindIII to have secondary reference for successful transformations. Predicted fragment around 1800 bp in size is seen in samples B1/5 and B1/6 (N.b. *PDS1*) and T1/1 and T1/4 (N.t. *PDS1*).

S2/4 (N.s. *PDS2*) and T2/2 (N.t. *PDS2*) gave different sized fragments compared to other PvuI/HindIII digested samples but have correct insert in them. This is due to HindIII cutting site in their respective *PDS2* fragments. These cutting sites result combined fragment (1834-1880 bp) splitting into two fragments in size of 1228 bp and 610 bp in S2/4 and 1434 bp and 410 bp in T2/2.

5.2.3 Sequencing results

PDS1 and *PDS2* gene fragment containing pJET1.2 plasmids were propagated in corresponding *E. coli* strains B1/5 and B1/6 (N.b. *PDS1*), B2/2 and B2/6 (N.b. *PDS2*), S1/1 and S1/2 (N.s. *PDS1*), S2/4 (N.s. *PDS2*), T1/1 and T1/4 (N.t. *PDS1*) and T2/2 (N.t. *PDS2*). Plasmids were then purified and send to sequencing.

Sequencing was done to compare whether genes of our own plants had differences caused by single-nucleotide polymorphism (SNP) compared to plants used to produce available public sequences of *PDS* genes. Sequence data from plasmids was compared using BLAST (National Center for Biotechnology Information, U.S. National Library of Medicine) against original genomic sequences of *PDS1* and *PDS2* used in primer and target sequence design phase. All plasmid sequences match with their assumed gene insert and primer binding sites. Gene insert was ligated into plasmid in forward orientation in all plasmids except in plasmids 4 and 9 where it was ligated in reverse orientation. *N. benthamiana PDS1* inserts in plasmids 7 and 8 had highest number of one base pair SNP differences and sequence gaps, followed by *N. benthamiana PDS2* inserts in plasmids 1 and 2, when sequences from university stock plants were compared to published *N. benthamiana PDS* sequence (Table 1).

N. sylvestris PDS1 insert in plasmids 3 and 4 and *PDS2* insert in plasmid 9 had low number of SNP differences and sequence gaps while *N. tomentosiformis PDS1* insert in plasmids 5 and 6 and *PDS2* insert in plasmid 10 had lowest number of SNP differences and sequence gaps when sequences from university stock plants were compared to published *N. sylvestris* and *N. tomentosiformis PDS* sequences. Deletion in sequence was only noted on *N. benthamiana PDS2* insert on plasmid 2 where significant part of 5' end sequence was missing including primer binding site and part of following sequence at total of 69 base pairs missing (Table 1).

HinfI target sites were identical between our plants and public sequences and did not contain any single-nucleotide polymorphisms thus any potential difference in efficiency of sgRNA binding to its target in plants is due to other factors than mutated HinfI site.

Table 1. Sequencing data from different plasmids. Plasmid sequences compared to *PDS1* and *PDS2* genomic sequences from *N. benthamiana* (Niben101Scf01283), *N. sylvestris* (NW_009521166) and *N. tomentosiformis* (NW_008898264) using Basic Local Alignment Search Tool (BLAST).

Gene insert	Plasmid number	Sequence orientation	1 bp SNP differences	Sequence gaps	Sequence deletions
NbPDS1	P. 7	Forward	31	6	-
NbPDS1	P. 8	Forward	33	6	-
NbPDS2	P. 1	Forward	22	2	-
NbPDS2	P. 2	Forward	15	2	Yes, 5'end
NsPDS1	P. 3	Forward	0	1	-
NsPDS1	P. 4	Reverse	0	1	-
NsPDS2	P. 9	Reverse	4	1	-
NtPDS1	P. 5	Forward	1	0	-
NtPDS1	P. 6	Forward	3	0	-
NtPDS2	P. 10	Forward	1	0	-

Comprehensive data from BLAST is presented in appendices for:
N. benthamiana PDS1 in plasmid no. 7 (Figure 18, appendices)
N. benthamiana PDS2 in plasmid no. 1 (Figure 19, appendices)
N. sylvestris PDS1 from *N. tabacum* in plasmid no. 3 (Figure 20, appendices)
N. sylvestris PDS2 from *N. tabacum* in plasmid no. 9 (Figure 21, appendices)
N. tomentosiformis PDS1 from *N. tabacum* in plasmid no. 5 (Figure 22, appendices)
N. tomentosiformis PDS2 from *N. tabacum* in plasmid no. 10 (Figure 23, appendices)

Plasmid sequences match with their assumed gene insert and primer binding sites indicating successful transformation of all different *PDS* loci into pJET1.2 plasmid. Gene insert was ligated into plasmid in forward orientation in all plasmids except in plasmids 4 and 9 where it was ligated in reverse orientation. Plasmid 4 is duplicate from plasmid 3 with NsPDS1 that has forward orientated insert. Plasmid 9 is only plasmid containing NsPDS2 thus this insert is only available as reverse orientated variant.

N. benthamiana inserts had highest number of single base pair differences and sequence gaps. This can be attributed to at least two different causes. First possible explanation is that there is significant difference in genetic background between university's *N. benthamiana* strain and source plant used in published genomic sequence of *N. benthamiana*. Secondly, in the planning stage of the experiment, *N. benthamiana* was erroneously thought to be diploid species that lead to situation where only one locus of both *PDS1* and *PDS2* was targeted. As *N. benthamiana* is a tetraploid species this means it is possible that the compared sequences of the genes represent different loci. This causes a situation where sequences match as right genes but have unusually high number of SNP differences due to being a different paralogous locus.

Minor sequence variation between university's *N. tabacum* strain and source plants used in published genomic sequences of *N. sylvestris* and *N. tomentosiformis* that together represent polyploid *N. tabacum*, indicate close genetic relationship segregated by few point mutations.

A deletion in sequence was only noted on *N. benthamiana PDS2* insert on plasmid 2 that had lost 69 bp fragment from 5' end.

HinfI cutting site functioning as Cas9 target site in *PDS2* is identical in genomes of both *N. tabacum* and *N. benthamiana*. *PDS1* in *N. tomentosiformis* genome of *N. tabacum* however is supposed to have one nucleotide difference in its HinfI cutting site compared to other species, but sequence data from our own plants showed that this difference does not exist in university's stock plants and the HinfI cutting site in *PDS1* of *N. tomentosiformis* is identical to the HinfI cutting site in *PDS1* of *N. sylvestris* genome of *N. tabacum* and *N. benthamiana*.

5.3 Rosetta and Lemo21 transformation

Two strains of *E. coli* were transformed with two different plasmids designed for Cas9 protein production to test which combination would be most efficient for Cas9 production.

When plasmid DNA is isolated from Rosetta and Lemo21 cells and digested with PstI/PvuII, successful transformants should show combination of digestion fragments originating from the transformant plasmid and the host organism's own plasmid.

Plasmid pET28a-Cas9-His does not have PstI cutting sites but has PvuII sites thus digestion with PstI/PvuII yields fragments in four sizes from PvuII digestion, which are 4753 bp, 3542 bp, 999 bp and 93 bp fragments.

Plasmid pET28b-Cas9-His has both PstI and PvuII cutting sites and digestion with PstI/PvuII yields fragments in sizes 2895 bp, 1968 bp, 1670, 1005 bp, 999 bp, 234 bp, 207 bp, 196 bp, 108 bp and 93 bp.

First transformation attempt (Figure 4A) yielded uniform plasmid digestion results from Rosetta(pET28a-Cas9-His), Lemo21(pET28a-Cas9-His) and Lemo21(pET28b-Cas9-His) transformants. Rosetta(pET28b-Cas9-His) transformants plasmid digestion results were incorrect as they were identical to Rosetta(pET28a-Cas9-His) results.

Second transformation attempt (Figure 4B) confirmed uniform results from first experiment but gave mixture of two different plasmid digestion results for Rosetta(pET28b-Cas9-His). Two digestions (RB1 and RB3) were identical to

Rosetta(pET28a-Cas9-His) results as in first experiment but third digestion (RB4) was unique and differing from Rosetta(pET28a-Cas9-His) indicating it was successful transformant.

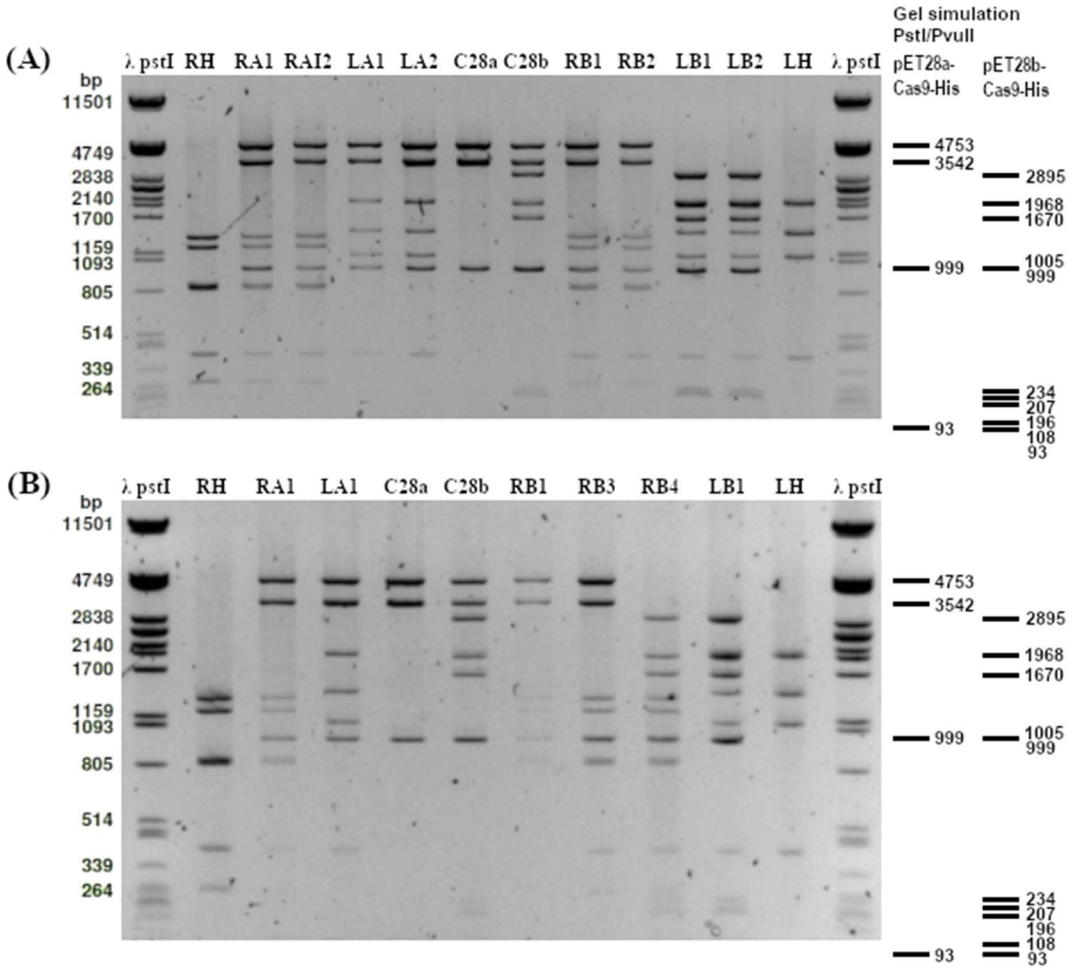


Figure 4. PstI/PvuII digested plasmids from *E. coli* strains Rosetta and Lemo21 transformed with pET28a-Cas9-His and pET28b-Cas9-His plasmids. Sample names are composed from R = Rosetta, L = Lemo21, A = pET28a-Cas9-His, B = pET28b-Cas9-His and number indicating selection plate sector from which colony was isolated. RH = Rosetta host and LH = Lemo21 host cell plasmids as controls. C28a and C28b are pET28a-Cas9-His and pET28b-Cas9-His plasmids as controls. **(A)** Digestion products from first transformation attempt. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. **(B)** Digestion products from second transformation attempt. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker.

5.3.1 Successfully transformed strains

Successful strains containing vector pET28a-Cas9-His were clearly recognized as combination of bands originating from both pET28a-Cas9-His and Rosetta host (RH) or

Lemo21 host (LH) plasmid. Successful strains were RA1, RA2, LA1 and LA2. One strain of each plasmid version was chosen and named for later use. These were RA1 which was named R(28a) and LA1 which was named L(28a).

Successful strains with pET28b-Cas9-His were different by not being full combination of bands from RH or LH plasmids and pET28b-Cas9-His but missing two largest bands (~5000 bp and ~3000 bp) from control pET28b-Cas9-His. This was likely due to original tube containing pET28b-Cas9-His being contaminated with pET28a-Cas9-His and two largest bands in fact originating from pET28a-Cas9-His thus transformation efficiency was random between both plasmids. This is seen in case of Rosetta from samples RB1 and RB2 (1st gel) and RB3 (2nd gel) that are transformed with pET28a-Cas9-His while only RB4 (2nd gel) has acquired pET28b-Cas9-His plasmid.

Plasmid pET28b-Cas9-His has both PstI and PvuII cutting sites and digestion with PstI/PvuII yields fragments in sizes 2893 bp, 1968 bp, 1672, 1007 bp, 999 bp and several less than 234 bp. This confirms contamination of pET28b-Cas9-His tube with pET28a-Cas9-His as actual digestion fragments of pET28b-Cas9-His are those seen gel simulation for pET28b-Cas9-His. Successful strains were LB1, LB2 and RB4. One strain of each plasmid version was chosen and named for later use. These were RB4 which was named R(28b) and LB1 which was named L(28b).

5.4 CAS9 production

Different production protocols for Cas9 protein were studied to discern which conditions would be most efficient for *E. coli* to grow and produce Cas9 protein.

5.4.1 Cas9 production results in protein gel

First experiment was conducted with strains R(28a), R(28b), L(28a) and L(28b) using induction temperatures of +22 °C (room temperature) for 4 hours and +37 °C for 2 hours as described in the methods. Preliminary results were received and used to decide which strains will be used in following experiments.

Second experiment was conducted only with R(28a) and R(28b) as Lemo21 based strains were deemed ineffective and were left out. Induction temperature of +28 °C for 2 hours was used and three different lysing products were compared in effectiveness: lysozyme, B-PER and CelLytic.

Third experiment (Figure 5) was conducted with R(28a) and R(28b) with induction temperature of +16 °C as recommended by Liang and colleagues for 3 hours and lysing with lysozyme (Liang et al. 2017).

Fourth experiment was conducted with strains R(28a) and R(28b) using induction temperature of +16 °C and induced overnight. In this experiment protein gel analysis was not done and main goal was to produce large amount of Cas9 for protein purification. 100 ml cultures with 40 µl of IPTG were grown from both R(28a) and R(28b).

Fifth experiment was conducted with strains R(28a) and R(28b) using induction temperatures of +16 °C overnight and +28 °C for 3 hours. Samples were lysed with B-PER products and after initial protein gel analysis second gel analysis was conducted with additional western blotting analysis.

Size of Cas9 is 162 kDa. Soluble protein samples of both R(28a) (Figure 5A) and R(28b) (Figure 5B) from control and induced samples show only faint band between 130 and 250 kDa apparently larger than 162 kDa. Total and insoluble protein samples from R(28a) and R(28b) control samples show a similar but stronger band larger than 162 kDa. Total and insoluble protein samples from R(28a) induced samples show same band as control samples. Total and insoluble protein samples from R(28b) induced samples show strong band closer in size to 162 kDa compared to other samples.

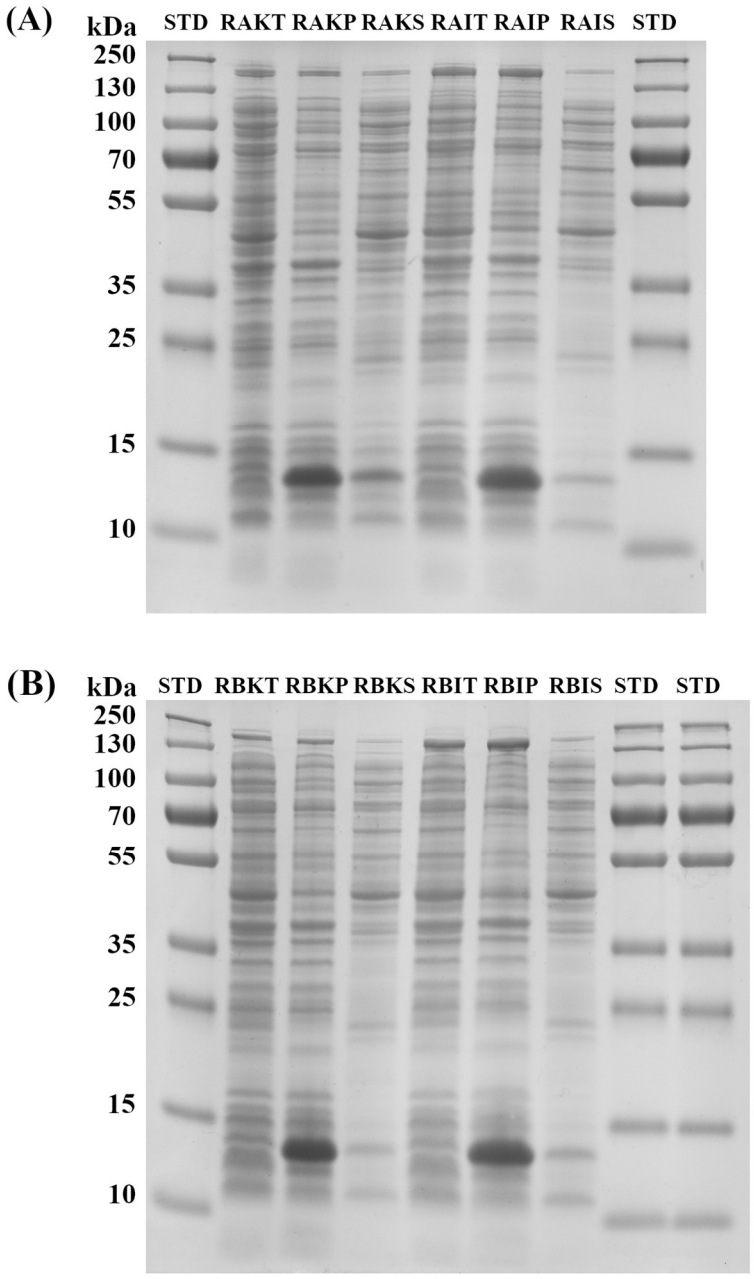


Figure 5. Extracted proteins from Cas9 production cultures. Sample names are composed from R = Rosetta, A = pET28a-Cas9-His, B = pET28b-Cas9, K = uninduced control, I = induced and final suffix of T = total protein, P = insoluble protein and S = soluble protein. **(A)** R(28a) samples from 3-hour cultures at +16 °C from third Cas9 production experiment. kDa, size of proteins as mass in kilodaltons determined by STD (standard protein ladder). **(B)** R(28b) samples from 3-hour cultures at +16 °C from third experiment. kDa, size of proteins as mass in kilodaltons determined by STD (standard protein ladder). Expected size of the expressed Cas9 protein is 162 kDa.

5.4.2 Protein purification results in protein gel

E. coli strains produce histidine-tagged Cas9 protein that enables protein purification to isolate target protein. His SpinTrap and PD MiniTrap G-25 (GE Healthcare) were used to purify produced protein to see in detail which size band correlates with our product.

Histidine-tagged protein of around 45 kDa in size was extracted from both R(28a) and R(28b) cultures originating from fourth Cas9 production experiment (Figure 6). This protein is not active functioning form of Cas9 that should be 162 kDa in size.

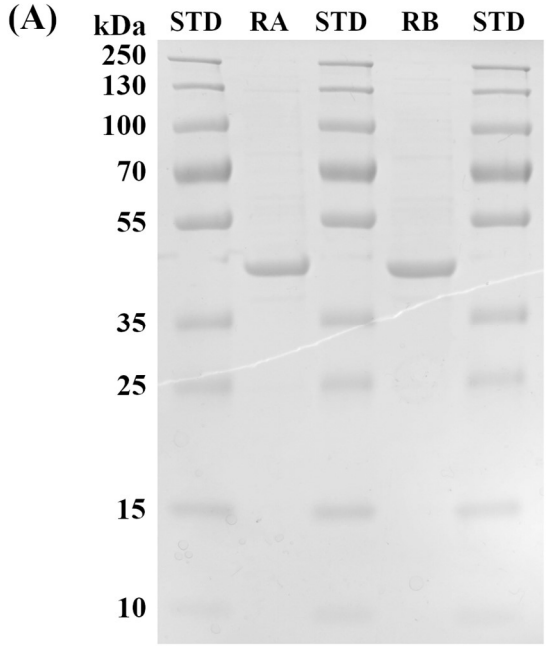


Figure 6. Purified proteins from Cas9 production cultures. Sample names are composed from R = Rosetta, A = pET28a-Cas9-His and B = pET28b-Cas9. **(A)** R(28a) and R(28b) samples from overnight cultures at +16 °C presenting extracted histidine-tagged proteins. kDa, size of proteins as mass in kilodaltons determined by STD (standard protein ladder).

5.4.3 Western blotting results

After negative results from protein purification, products from *E. coli* were studied using western blotting to see full picture of all histidine-tagged proteins in protein gel. Western blotting was done for protein gel originating from fifth Cas9 production experiment.

Light signal caused by chemiluminescence from secondary antibody bound to histidine-tagged proteins is seen in induced total and insoluble protein samples of both R(28a) and R(28b) but not in soluble protein samples. Figure 7A shows cultures grown at +16 °C. R(28a) samples show bands in approximate sizes of 162 kDa (Cas9) and 50 kDa. R(28b) samples show bands in approximate sizes of 162 kDa (Cas9), 45 kDa and 35 kDa.

Figure 7B shows cultures grown at +28 °C. R(28a) samples show bands in approximate sizes of 162 kDa (Cas9), 50 kDa and 24 kDa. R(28b) samples show one band in approximate size of 45 kDa.

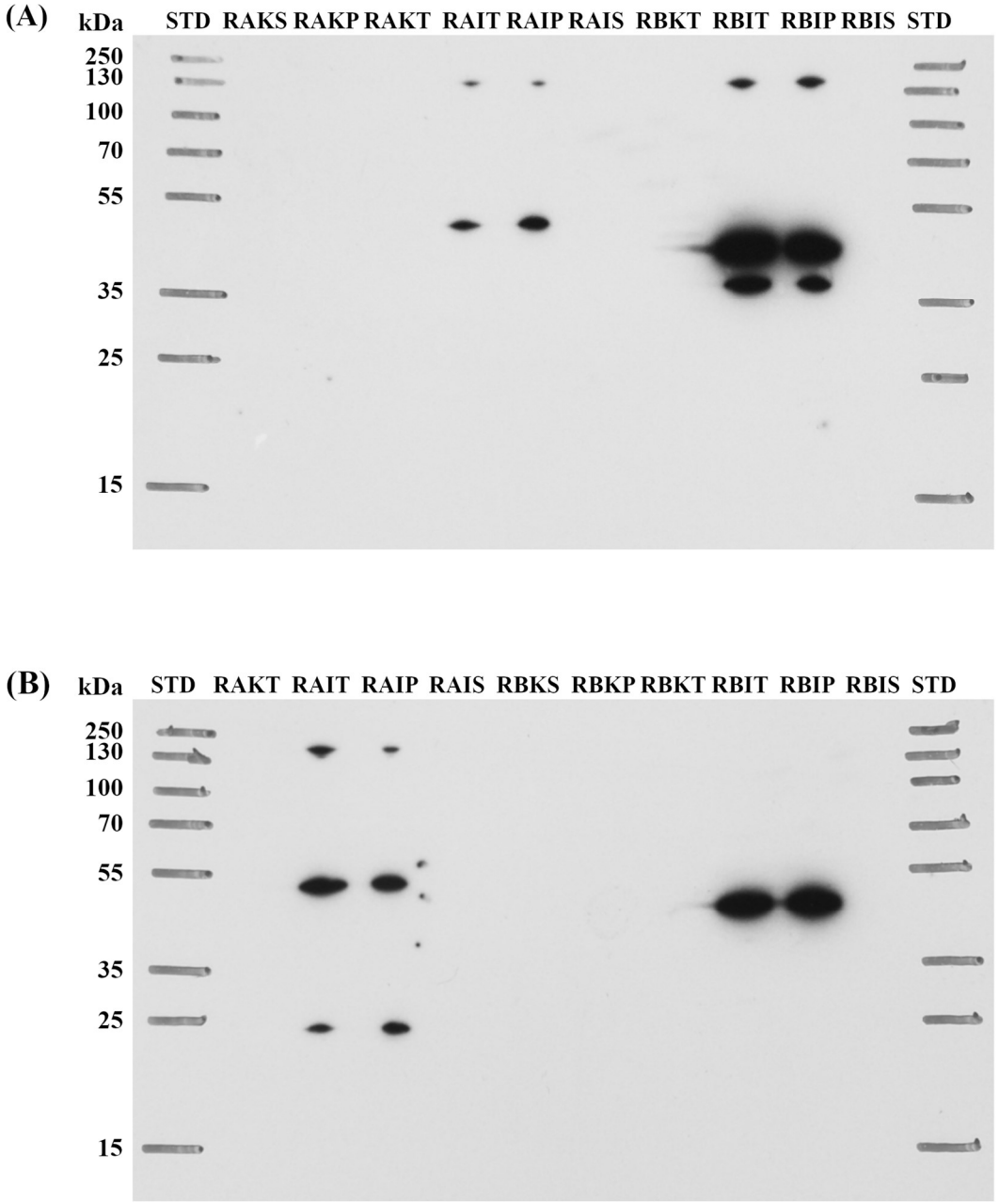


Figure 7. Western blotting results of extracted proteins from Cas9 production cultures. X-ray film exposed for 30 minutes. Sample names are composed from R = Rosetta, A = pET28a-Cas9-His, B = pET28b-Cas9, K = uninduced control, I = induced and final suffix of T = total protein, P = insoluble protein and S = soluble protein. **(A)** R(28a) and R(28b) samples from overnight cultures at +16 °C. kDa, size of proteins as mass in kilodaltons determined by STD (standard protein ladder). **(B)** R(28a) and R(28b) samples from 3-hour cultures at +28 °C. kDa, size of proteins as mass in kilodaltons determined by STD (standard protein ladder).

5.4.4 Conclusions from Cas9 production results

Protein gel analysis showed strong bands between 130 kDa and 250 kDa in size from induced samples but due to other natural proteins in same size category from *E. coli*, results were too inaccurate to proclaim whether one of them was Cas9 protein. Western blotting confirmed presence of 162 kDa Cas9 protein in gel area between 130 kDa and 250 kDa but also confirmed that Cas9 was only found from insoluble fractions and all samples containing Cas9 also had degraded versions of Cas9 proteins in different sizes. One of these degraded Cas9 proteins was extracted from soluble fraction and was showing as 45 kDa histidine-tagged protein in gel.

Lower induction temperature of +16 °C was suggested by earlier studies (Liang et al. 2017) and while no soluble Cas9 protein was produced, +16 °C was better induction temperature than +28 °C as in higher temperature R(28b) produced only degraded Cas9 proteins. While R(28a) produced right sized 162 kDa Cas9 protein also at +28 °C induction temperature, it also produced smallest degraded 24 kDa Cas9 protein that was not seen at +16 °C, thus supporting lower induction temperature as better choice. Study by Ramakrishna diverts from low induction temperature as Cas9 protein was successfully produced overnight at +30 °C (Ramakrishna et al. 2014). All in all, the Cas9 production experiment was not successful with methods used in this experiment.

5.4.5 Reasons for insolubility

Manual for pET system mentions that in case of hydrophobic or Membrane-Associated Domain Proteins, soluble proteins can precipitate into insoluble fraction if lysing buffer contains NaCl and lysozyme is most dependable lysing agent. Lysozyme was used during early protein gel runs but in later experiments in which soluble histidine-tagged proteins were extracted and in last experiment conducting western blotting analysis, B-PER product was used in protein extraction. It is thus possible that if protein extraction and western blotting were done from samples lysed with lysozyme, evidence of soluble Cas9 protein could have been achieved and failure to produce soluble protein was due to poor choice of lysing agent.

Previous studies have used various versions of lysing methods. In study by Liang there is no mention of lysing method (Liang et al. 2017). Liu used sonification for lysing and buffer containing 500 mM NaCl but produced soluble Cas9 protein (Liu et al. 2015). Ramakrishna used sonification for lysing and buffer containing 300 mM NaCl and 1 mg/ml lysozyme (Ramakrishna et al. 2014).

5.5 Agroinfiltration

Transient transformation with CRISPR-Cas9 was studied using agroinfiltration to *N. benthamiana* leaves to test functionality and efficiency of the designed *Agrobacterium* strains before attempting stable transformation through full agrotransformation. Successful mutations where analyzed with two different mutation detection methods, which were commercial T7E1 mutation detection kit and celery extract containing natural CEL I endonuclease.

5.5.1 Expected digestion fragments

When DNA samples from agroinfiltrated leaves are treated with endonuclease suitable for mutation detection it cuts mismatch bubbles in heteroduplex between non-mutated and mutated PCR products. As this bubble is cut by endonuclease it shows two distinct bands in gel revealing successful mutations.

CRISPR-Cas9 transformation done by single transcription unit cuts DNA from three base pairs upstream from the PAM sequence in target sequence. Mutation detection from *N. benthamiana PDS1* PCR fragment (622 bp) should yield two fragments from mutated sequence in length of 64 bp and 558 bp. Gel images should thus show two bands from which higher band represents 622 bp fragments from non-mutated cells while second band shows digested larger 558 bp fragments from successfully mutated cells.

Mutation detection from *N. benthamiana PDS2* PCR fragment (657 bp) should yield two fragments from mutated sequence in length of 191 bp and 466 bp, which in gel should show 657 bp and 466 bp bands in case of successful mutation.

T7E1 mutation detection kit comes with control template and primer mix for producing heteroduplex that has digestion products of 200 bp and 400 bp.

An additional control was made by amplifying the inserts from plasmids pHTT837 and pHTT838 and generating heteroduplexes with two mismatch bubbles.

5.5.2 Mutation detection with T7E1

Mutation detection was done for DNA samples from agroinfiltrated leaves by T7 Endonuclease 1 assay.

First agroinfiltration experiment was done with water control and the still unselected pDAR1 and pDAR2 carrying *Agrobacterium* strains E1-2A, E1-2B, E1-3A, E1-3B, E2-2A, E2-2B, E2-3A and E2-3B. Leaf samples were collected three days after agroinfiltration. Samples were analyzed using HinfI digestion. Samples from plants infiltrated with E1-2A, E1-2B, E1-3A and E1-3B that represent *PDS1* showed successful mutation based HinfI digestion but only in PCR products made with *PDS2* primers. E2-2A, E2-2B, E2-3A and E2-3B samples representing *PDS2* showed correct digestion product only from E2-2A. Despite ambiguous results possibly resulting from sample mix-up, E1-2A was chosen and named as DAT1 while E2-2A was named DAT2.

Second experiment was done with water control and *Agrobacterium* strains DAT1 (E1-2A) and DAT2 (E2-2A) selected from previous experiment representing each group. Infiltration combination with strains DAT1 + P19 and DAT2 + P19 was also tested to see whether *Agrobacterium* strain P19 enhances expression rate. Leaf samples were collected in days 1, 2, 3 and 6.

DAT1 samples gave pronounced second band in samples digested with T7/E1 (Figure 9A). Higher band represents 622 bp fragments from non-mutated cells or non-digested mutated fragments while second band shows digested larger 553 bp fragments from successfully mutated cells. DAT2 samples did not show clear second band indicating lower efficiency in strains ability to introduce mutations. Control PCR products and PCR products from water infiltrated plants show only one band with no additional second band. Sample from plant infiltrated with both DAT1 and P19 show darker band compared to

sample from plant infiltrated only with DAT1. Control reactions with plasmid from T7E1 kit show similar results.

DAT1 samples from all collection days (1, 2, 3 and 6) after infiltration were analyzed on second image (Figure 9B). Second band is strongest on samples collected 6 days after infiltration with lower concentration on samples collected 1, 2 or 3 days after infiltration, particularly in samples from plants infiltrated with both DAT1 and P19. Samples from plants infiltrated only with DAT1 show more gradual increase in second band concentration from day 1 to day 6.

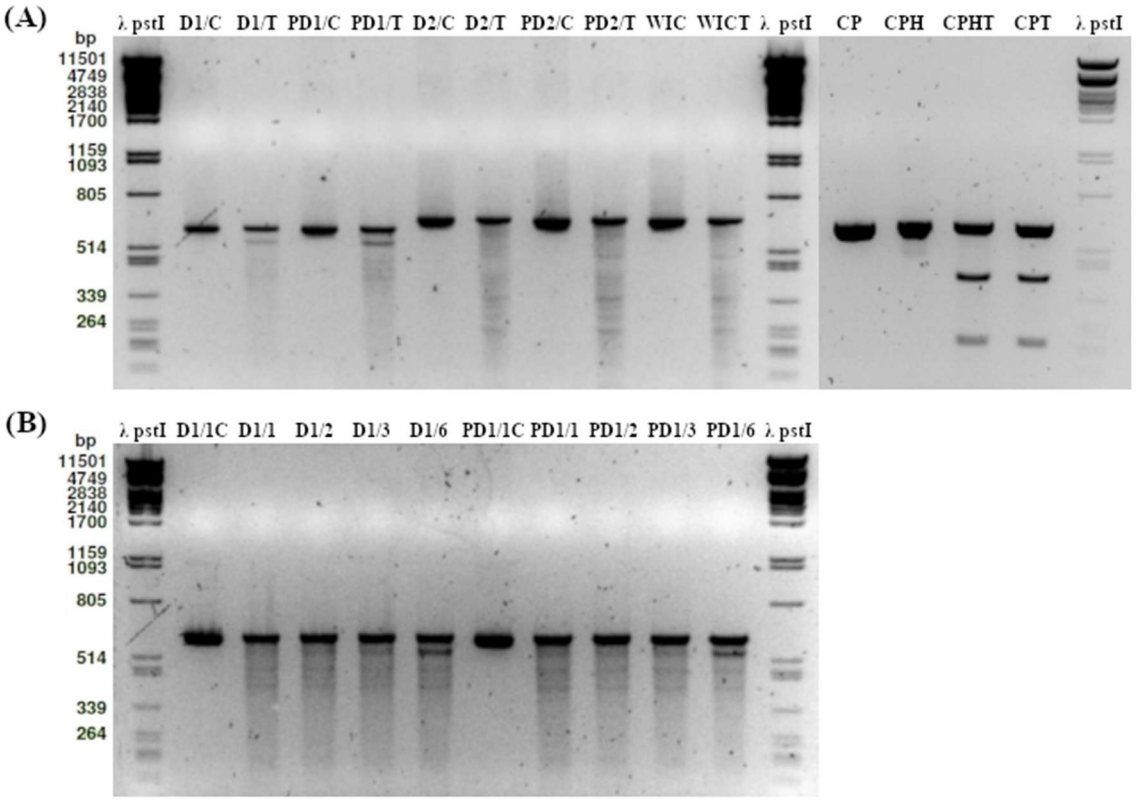


Figure 9. PCR products from second *N. benthamiana* agroinfiltration experiment digested with T7E1. Sample names are composed from D1 = DAT1, D2 = DAT2, PD1 = P19 + DAT1, PD2 = P19 + DAT2, C = undigested control PCR sample and T = T7E1 digested sample. WIC = water infiltrated control, WICT = WIC digested with T7E1 as control plants without *Agrobacterium*, PCR products made with *PDS2* primers. **(A)** Digestion products from second agroinfiltration experiment with samples collected 6 days after infiltration. Additional samples on right side are PCR products from T7E1 kit's control plasmid. CP = control PCR product, CPH = control PCR product after heteroduplex formation, CPHT = CPH digested with T7E1 and CPT = CP digested with T7E1. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. **(B)** Digestion products from second agroinfiltration experiment with DAT1 samples collected 1, 2, 3 and 6 days after infiltration indicated by /1, /2, /3 and /6 suffixes in sample names. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. See text for expected fragment sizes.

Third experiment was done to verify results. It was direct replication of second experiment and yielded equal results thus verifying earlier results. Gel images from second experiment were chosen as representatives of the experiments as they were of better quality due to stronger staining and image resolution.

T7E1 mutation detection was effective and gave clear results when used but second band indicating successful transformation was not as strong compared to bands from kit's control reaction and long exposure was needed for agroinfiltration samples. Strongest second band marking successful mutation was in samples collected 6 days after

infiltration and difference between plants infiltrated either with only DAT1 or with mixture of both DAT1 and P19 was minor, indicating post-transcriptional silencing was low in plants, and suppressing effect of P19 infiltration was not significant compared to infiltration with only DAT1 without suppressing effect.

5.5.3 Mutation detection with celery extract

Alternative mutation detection method with natural CEL I endonuclease was studied by producing this endonuclease from fresh celery and comparing its mismatch bubble cutting efficiency to commercial T7E1.

Celery extract experimenting started from cutting of controlled heteroduplex formed between PCR products of pHTT837 and pHTT838 plasmids (Figure 17, appendices). These two plasmids are similar except one changed restriction enzyme recognition site. pHTT837 has HindIII site in position 826 while pHTT838 has EcoRI in position 1172. Mismatches are formed in these positions when the PCR products are combined into heteroduplex. Mutation detection from this fragment should yield six different cutting outcomes that are 727 bp (undigested PCR product), 631 bp (only EcoRI bubble cut), 553 bp (only HindIII bubble cut), 346 bp (both bubbles cut), 207 bp (tail after cut EcoRI) and 174 bp (tail before cut HindIII).

Digestion of mismatch bubble formed from pHTT837 and pHTT838 plasmids PCR products combined into heteroduplex gave strong, easily interpreted results from both digestion methods of either T7E1 or celery extract with several clear bands compared to single band of control PCR product. T7E1 serial dilution's digestion bands smaller than ~350 bp were more concentrated in strongest dilution compared to ~600 bp band while in weakest dilution opposite was true and ~600 bp band was more concentrated compared to bands smaller than ~350 bp. In serial dilution samples digested with celery extract both bands smaller than ~350 bp and ~600 bp band weakened in concentration equally from strongest to weakest dilution (Figure 10).

Crude celery extract without DNA in gel (CC) and samples containing DNA digested with celery extract show blurry noise at ~500-800 bp and >100 bp region of the gel. Celery extract treated with RNase (CR) did not show noise at >100 bp region of the gel

and noise at ~500-800 bp region of the gel was highest at sample with most RNase added and slightly less in RNase serial dilution's weakest sample. Celery extract washed with ethyl acetate in order degrade DNA contaminants from the sample to had noise only at >100 bp. Celery extract washed with both ethyl acetate and acetic acid shows almost empty line with minor noise at >100 bp region of the gel (Figure 10).

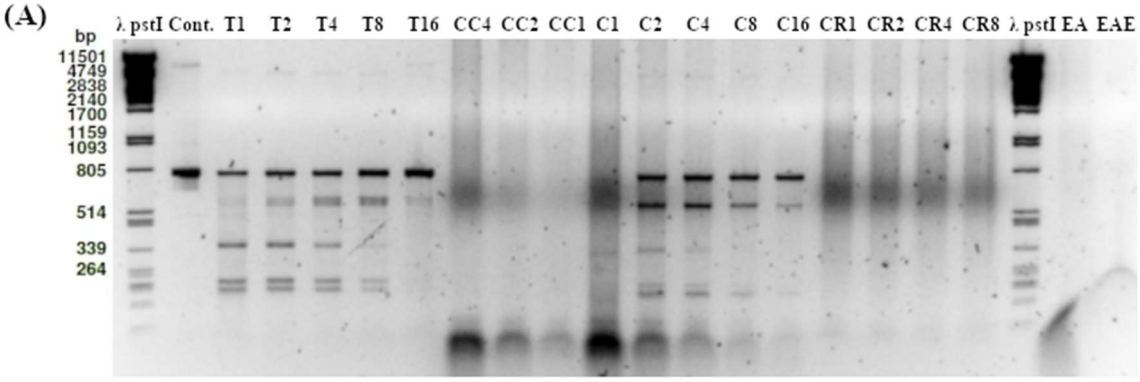


Figure 10. Combined heteroduplex PCR products from pHTT837 and pHTT838 plasmids digested with T7E1 or celery extract in form of serial dilutions and pure celery extract processed with various methods. Sample names are composed from T1 to T16 = T7E1 serial dilution with DNA from 1 µl to 0.0625 µl, C1 to C16 = celery extract serial dilution with DNA from 1 µl to 0.0625 µl, CC4 to CC1 = crude celery extract 4, 2 and 1 µl (without DNA), CR1 to CR8 = 4 µl of celery extract with RNase serial dilution from 1 µl to 0.125 µl (without DNA), EA = celery extract washed with ethyl acetate and EAE = acetic acid acidified celery extract washed with ethyl acetate. Cont. = heteroduplex PCR product as control. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker.

Digestion of agroinfiltration samples from third experiment with celery extract gave similar results to digestion with T7E1 in second experiment but with lower fidelity. Samples from DAT1 infiltrated plants showed detectable second band but in samples from DAT2 infiltrated plants no clear second band formed (Figure 11A). PCR products from water infiltrated plants show only one band with no additional second band. Sample from plant infiltrated with both DAT1 and P19 show second band equal in concentration with DAT1.

DAT1 samples from all collection days (1, 2, 3 and 6) after infiltration were analyzed with celery extract on second image (Figure 11B). Second band is strongest on DAT1 sample collected 6 days after infiltration with lower concentration on samples collected 1, 2 or 3 days after infiltration. Samples from plants infiltrated with both DAT1 and P19 show weak second band with equal concentration from all collection days.

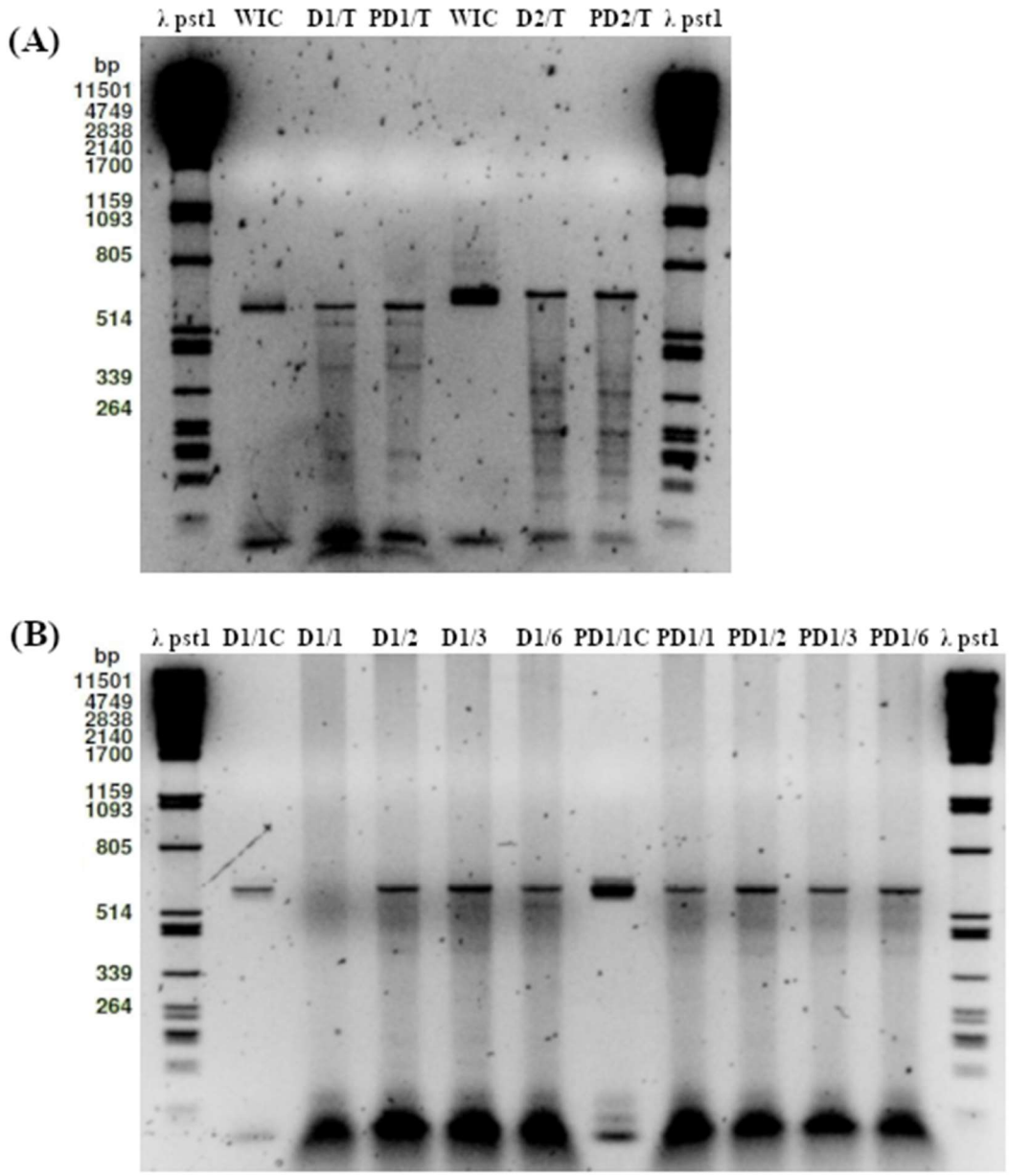


Figure 11. PCR products from third *N. benthamiana* agroinfiltration experiment digested with celery extract. Sample names are composed from D1 = DAT1, D2 = DAT2, PD1 = P19 + DAT1, PD2 = P19 + DAT2, T = celery extract digested sample, WIC = water infiltrated control and C = control PCR sample. **(A)** Digestion products from third agroinfiltration experiment with samples collected 6 days after infiltration. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker. **(B)** Digestion products from third agroinfiltration experiment with DAT1 samples collected 1, 2, 3 and 6 days after infiltration indicated by /1, /2, /3 and /6 suffixes in sample names. Bp, size of digested fragments as length in base pairs determined by λ PstI size marker.

Contaminants in celery extract shown as blurry noise at ~500-800 bp and >100 bp region of the gel are likely two different compounds. Celery extract treated with RNase did not show noise at lengths less than 100 bp indicating these were contaminant RNA from celery. Noise at ~500-800 bp region of the gel did not disappear following RNase treatment indicating other contaminant than RNA but was dissipated after ethyl acetate

and/or acetic acid wash. Ethyl acetate wash was used to remove small molecules causing fluorescens such as coumarin and removal of noise at ~500-800 bp region of the gel indicates it was likely resulting from this source. Further adding of acetic acid was done to protonate weak acids such as phenolic compounds in samples and thus aid their infusion into ethyl acetate. This addition removed almost all detectable compounds from samples. Till considers crude extracts to be as good as purified CEL I in function but from the original studies figures it is clear that their crude extract had similar background noise, which must be considered when results are interpreted (Till et al. 2004).

Adding excessive amount of celery extract in samples resulted complete disintegration of DNA and significantly decreased resolution. This is due to crude extract containing several different enzymes in addition to desired CEL I enzyme including DNases and other disruptive compounds that disintegrate samples if relative concentration to sample DNA is too high. Till observed same effects and concluded that too high concentration of CEL I causes too many background bands and as mismatch products become concentrated the detection rate diminishes (Till et al. 2004).

Because of limited volume that could be used in samples, celery extract mismatch cleaving was not as effective as using concentrated T7E1 and gave fainter second band. When relevant results were first discovered using T7E1, celery extract could be used to verify results as even fainter digestion bands could be deemed correct when T7E1 reference was available. Analyzing mutation results using only celery extract is likely problematic and leaves ambiguity, which can lead to wrong interpretations but can function as economical solution for verifying results and analyzing replication samples.

5.6 Agrotransformation

Stable transformation with CRISPR-Cas9 was studied using agrotransformation to *N. tabacum* leaves. Successful transformation was studied by observing leaf color for possible photobleaching phenotype indicating mutation in *PDS* genes.

5.6.1 Transformation with DAT1 and DAT2

In first experiment two strains of *Agrobacterium* were tested, which were DAT1 and DAT2 with two replicates on transform plates with only claforan and on transformation

selection plates with claforan and hygromycin (50 mg/l). Non-agrotransformed control plates were 0-plate without antibiotics, Claf-plate with only claforan and Claf-Hygr-plate with claforan and hygromycin. After shoots grew too large to fit on petri dish, leaf pieces were moved to glass jars with same MS-media as in their corresponding plates for continued observation.

Leaf pieces on 0-plate without antibiotics grew well in healthy green color. Leaf pieces on regeneration control plate with only claforan grew well in healthy green color. Leaf pieces on regeneration control plate with both claforan and hygromycin died eventually without growing shoots. Leaf pieces transformed with DAT1 and DAT2 on plates containing only claforan grew promising white callus, but only green shoots emerged that eventually grew into large size on media jars but showed no sign of photobleaching. Leaf pieces transformed with DAT1 and DAT2 on selection plates containing claforan and hygromycin died without growing shoots.

Transformation with original *Agrobacterium* strains DAT1 and DAT2 resulted in failure that was likely due to using hygromycin as selection antibiotic. Hygromycin is feasible choice for selection but is considered as strong selection agent with narrow concentration window that can easily result in failure due to unsuitable concentration or general sensitivity of transformant. Instead of optimizing hygromycin concentration further, selection marker was changed to kanamycin due to its wider concentration window and general ease of use.

5.6.2 Transformation with POG1 and POG2

Second experiment was done with different *Agrobacterium* strains, which were POG1 and POG2 that have resistance to kanamycin instead of hygromycin thus kanamycin (100 mg/l) was used on selection plates. Non-agrotransformed control plates were Claf-plate with only claforan and Claf-Kana-Plate with claforan and kanamycin. Transformation plates contained claforan and kanamycin. Shoot growth was observed only on petri dish stage and leaf pieces were not grown further on glass jars.

Leaf pieces on regeneration control plate (Figure 12A) with only claforan grew well in healthy green color. Leaf pieces on regeneration control plate with both claforan and

kanamycin (Figure 12B) died eventually without growing shoots. Leaf pieces transformed with POG1 on selection plate containing claforan and kanamycin (Figure 12C) grew yellowish green shoots and one shoot with white-edged leaves. Leaf pieces transformed with POG2 on selection plate containing claforan and kanamycin (Figure 12D) grew white callus and one piece had green shoots together with two white shoots. POG2 plate had problem with *Agrobacterium* overgrowth, which destroyed few promising leaf pieces with white callus.

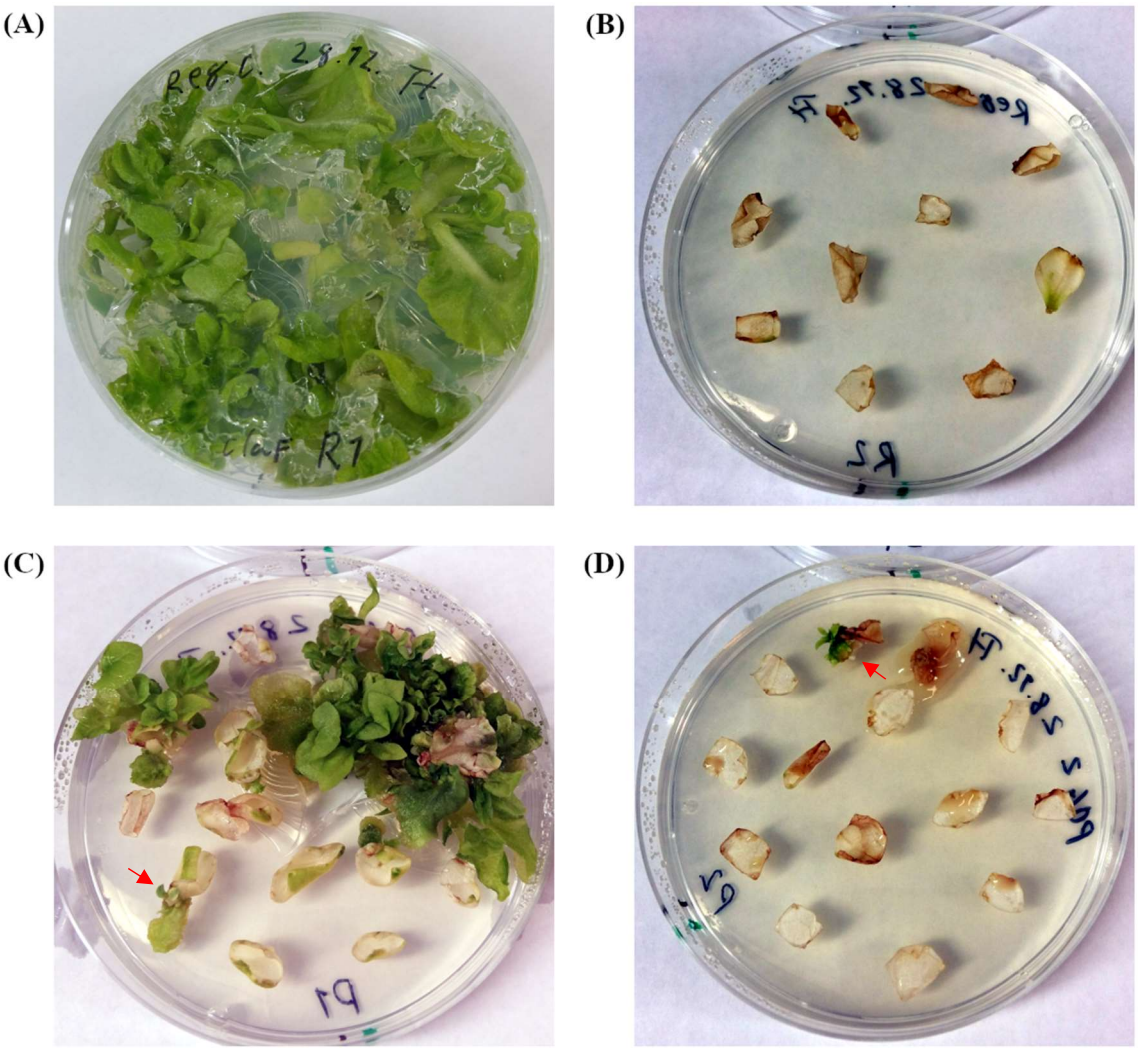


Figure 12. MS media plates with *N. tabacum* leaf pieces. **(A)** Regeneration control plate with claforan. **(B)** Regeneration control with claforan and kanamycin. **(C)** Selection plate containing claforan and kanamycin with leaf pieces transformed with POG1. Arrow pointing shoot with white-edged leaves **(D)** Selection plate containing claforan and kanamycin with leaf pieces transformed with POG2. Arrow pointing leaf piece with white shoots.

Kanamycin was more gentle selection agent compared to hygromycin and allowed sufficiently long survival of POG1 and POG2 transformed leaf pieces on selection plates for shoot growth to occur. POG1 transformed leaf pieces grew yellowish green shoots (Figure 12C) and one shoot with white edges showing expected photobleaching phenotype (See close-up in Figure 13A). POG2 transformed leaf pieces produced white shoots indicating full albino plants (See close-up in Figure 13B). Although these *Agrobacterium* strains could transform only either *PDS1* or *PDS2* locus ruling out full tetra-allelic mutations, it has been reported that albino and photobleaching phenotypes can occur in plants with partial multi-allelic mutations (Hooghvorst et al. 2019).

Allotetraploid genome of *N. tabacum* is unstable and contains several loss-of-function alleles (Chen et al. 2018) and homeolog silencing in general (Clarkson et al. 2017). Albino and photobleaching genotypes that are expressed without full tetra-allelic mutations could be hypothesized to results from partial or full silencing of some *PDS* loci in *N. tabacum,* but further studies should be made to confirm whether this is the cause.

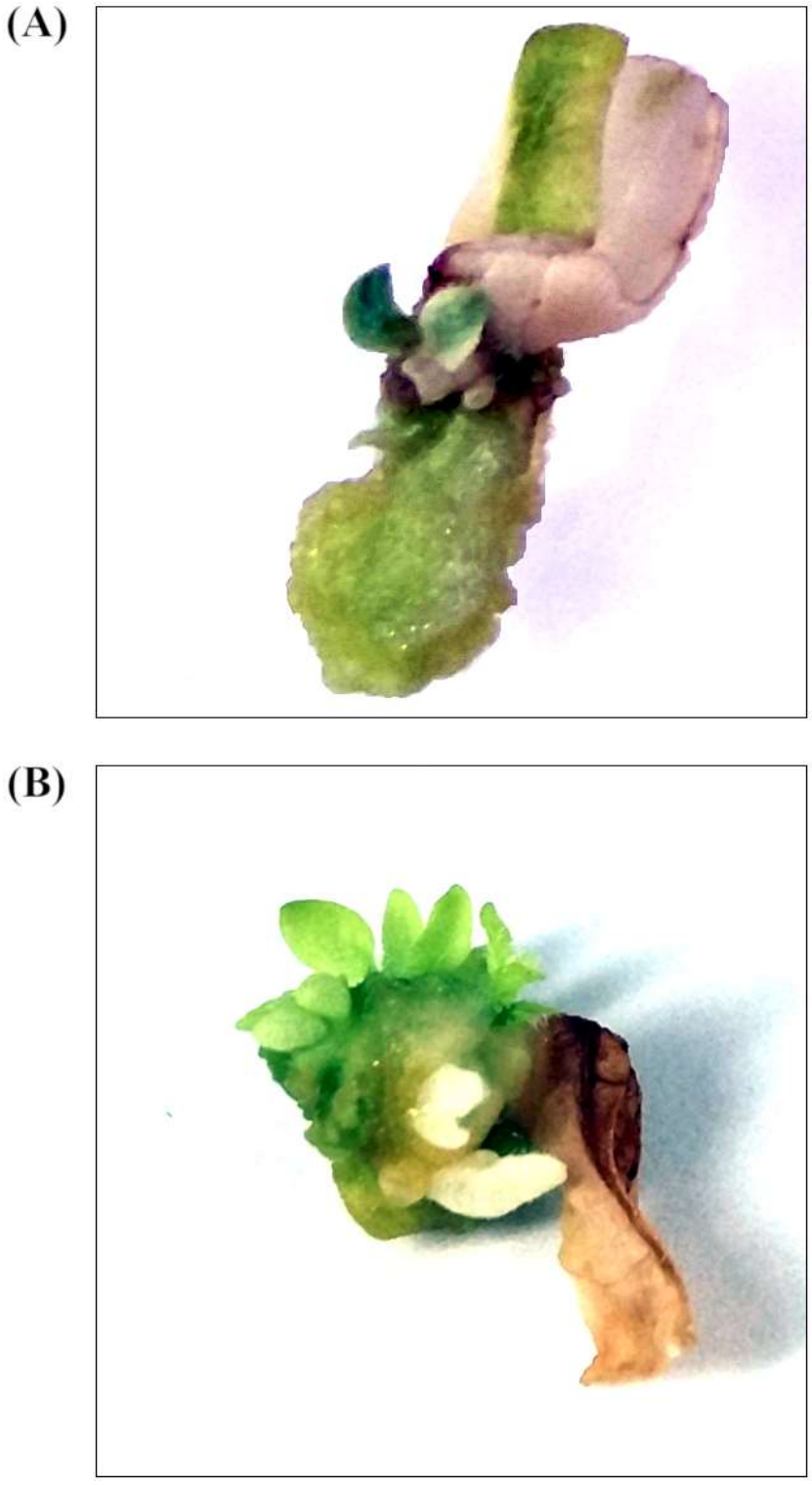


Figure 13. Close-up of leaf pieces showing shoots with photobleaching phenotype **(A)** POG1 transformed leaf piece showing shoot with white-edged leaves **(B)** POG2 transformed leaf piece showing white shoots.

6 CONCLUSIONS

In this master's thesis two different methods of CRISPR-Cas9 induced gene editing in the genus *Nicotiana* were compared together with two different methods for mutation detection.

CRISPR-Cas9 ribonucleoprotein complex mediated transformation was supposedly straightforward process from Cas9 protein production to assembling the complex and transforming cells. However, experiments never reached the transformation phase as producing active form of Cas9 protein in soluble form was challenging and this was not realized despite using different *E. coli* strains and production conditions. Transforming *E. coli* with preassembled CRISPR-Cas9 production vectors was easy and mass production of Cas9 is feasible idea in large scale experimenting, but this study suggests that acquiring ready-to-use Cas9 protein might be preferable choice when targeting only few transformations with CRISPR-Cas9 RNP-complex.

Agrotransformation with Single Transcriptional Unit CRISPR-Cas9 system was significantly more straightforward in practice after intricate plasmid design was realized and successfully transformed into *Agrobacterium* strains as CRISPR-Cas9 transformation was then simply produced naturally and effectively by bacterial cells introduced into plant cells. This method however lacks the several advantages that using CRISPR-Cas9 RNP-complex has such as instant gene editing in cell, avoiding RNA interference and transformation over species boundaries that species limited agrotransformation cannot achieve. Agrotransformation is well established method for genus *Nicotiana* and can be recommended when working with CRISPR-Cas9. Possible off-target mutations can be selectively outbred from successful transformants by crossbreeding if necessary.

Detecting successful mutations was easy with commercial T7E1 but interestingly mutation detection with natural CEL I endonuclease from celery extract was deemed very effective although contaminants in celery extract cause minor blurry noise that weakens band resolution in gel. Celery extract can be used at least as cost-effective alternative to verify or do replication for previous confirmed results from T7E1 mutation detection.

7 ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Teemu Teeri for providing me the thesis subject and support through the entire project in a subject that was new to our department at the time of beginning the experiments.

I would like to express gratitude for visiting exchange students Daria Pajak and Paloma Ortiz García who produced some of the key components used in the experiments.

I would also like to thank all people working in our lab and giving me support in using lab machinery and calculations in times of difficulty.

Finally, many thanks to my family for supporting me through the writing process.

REFERENCES

Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., Ding, S. & Mahfouz, M. 2018. RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biology* 19: 1.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A. & Liu, D. R. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576: 149-157.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-1712.

Bibikova, M., Beumer, K., Trautman, J. & Carroll, D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science* 300: 764-764.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. & Bonas, U. 2009. Breaking the code of DNA binding specificity of TAL-Type III effectors. *Science* 326: 1509-1512.

Bolotin, A., Ouinquis, B., Sorokin, A. & Ehrlich, S. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151: 2551-2561.

Chen, S., Ren, F., Zhang, L., Liu, Y., Chen, X., Li, Y., Zhang, L., Zhu, B., Zeng, P., Li, Z., Larkin, R .M. & Kuang, H. 2018. Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. *Molecular Plant* 11: 914-927.

Cho, S. W., Lee, J., Carroll, D., Kim, J. & Lee, J. 2013. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* 195: 1177-1180.

Clarkson, J. J., Dodsworth, S. & Chase, M. W. 2017. Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (*Solanaceae*). *Plant Systematics and Evolution* 303: 1001-1012.

Cohen, S. N. & Chang, A. C. Y. 1973. Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants. *Proceedings of the National Academy of Sciences of the United States of America* 70: 1293-1297.

Deblaere, R., Bytebier, B., Degreve, H., Deboeck, F., Schell, J., Vanmontagu, M. & Leemans, J. 1985. Efficient octopine Ti plasmid-derived vectors for agrobacterium-mediated gene transfer to plants. *Nucleic Acids Research* 13: 4777-4788.

Dellaporta, S. L., Wood, J. & Hicks, J. B. 1983. A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* 1: 19-21.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J., Shi, Y. & Yan, N. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335: 720-723.

Epinat, J., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. & Lacroix, E. 2003. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research* 31: 2952-2962.

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C. & Schier, A. F. 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *Plos One* 9: e98186.

Gallego-Bartolome, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H. Y., Zhao, J. M., Segal, D. J. & Jacobsen, S. E. 2018. Targeted DNA demethylation of the *Arabidopsis* genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2125-E2134.

Gao, H., Smith, J., Yang, M., Jones, S., Djukanovic, V., Nicholson, M. G., West, A., Bidney, D., Falco, S. C., Jantz, D. & Lyznik, L. A. 2010. Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant Journal* 61: 176-187.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P. & Xia, Q. 2015. CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Molecular Biology* 87: 99-110.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I. & Liu, D. R. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551: 464-471.

Grizot, S., Epinat, J., Thomas, S., Duclert, A., Rolland, S., Paques, F. & Duchateau, P. 2010. Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Research* 38: 2006-2018.

Haseloff, J. & Gerlach, W. L. 1988. Simple RNA enzymes with new and highly specific endoribonuclease activities. *Nature* 334: 585-591.

Hensgens, L. A., Bonen, L., de Haan, M., van der Horst, G. & Grivell, L. A. 1983. Two intron sequences in yeast mitochondrial *COX1* gene - homology among URF-containing introns and strain-dependent variation in flanking exons. *Cell* 32: 379-389.

Hiatt, W. R. 1994. Commercial introduction of Flavr Savr(tm) tomatoes. *Journal of Cellular Biochemistry*: 74-74.

Hooghvorst, I., Lopez-Cristoffanini, C. & Nogues, S. 2019. Efficient knockout of phytoene desaturase gene using CRISPR/Cas9 in melon. *Scientific Reports* 9: 17077.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z. & Liu, D. R. 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556: 57-63.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. & Zhang, B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nature Biotechnology* 29: 699-700.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. 1987. Nucleotide-sequence of the *iap* gene, responsible for alkaline-phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene-product. *Journal of Bacteriology* 169: 5429-5433.

Jackson, D. A., Berg, P. & Symons, R. H. 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40 - circular SV40 DNA molecules containing lambda phage genes and galactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 69: 2904-2909.

Jacquier, A. & Dujon, B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 41: 383-394.

Jansen, R., van Embden, J., Gaastra, W. & Schouls, L. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* 43: 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821.

Jurica, N., Monnat, R. & Stoddard, B. 1998. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Molecular Cell* 2: 469-476.

Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318: 648-651.

Khan, M. Z., Haider, S., Mansoor, S. & Amin, I. 2019. Targeting plant ssDNA viruses with engineered miniature CRISPR-Cas14a. *Trends in Biotechnology* 37: 800-804.

Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research* 24: 1012-1019.

Kim, Y., Cha, J. & Chandrasegaran, S. 1996. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* 93: 1156-1160.

Klein, T. M., Wolf, E. D., Wu, R. & Sanford, J. C. 1987. High-velocity microprojectiles for delivering nucleic-acids into living cells. *Nature* 327: 70-73.

Klug, A. & Rhodes, D. 1987. Zinc fingers - a novel protein fold for nucleic-acid recognition. *Cold Spring Harbor Symposia on Quantitative Biology* 52: 473-482.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533: 420-424.

Kostriken, R., Strathern, J., Klar, A., Hicks, J. & Heffron, F. 1983. A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell* 35: 167-174.

Li, J., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G. M. & Sheen, J. 2013. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology* 31: 688-691.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y. & Gao, C. 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications* 8: 14261.

Liu, J., Nannas, N. J., Fu, F., Shi, J., Aspinwall, B., Parrott, W. A. & Dawe, R. K. 2019. Genome-scale sequence disruption following biolistic transformation in rice and maize. *Plant Cell* 31: 368-383.

Losey, J. E., Rayor, L. S. & Carter, M. E. 1999. Transgenic pollen harms monarch larvae. *Nature* 399: 214-214.

Luo, M., Pang, C., Gerken, A. & Brock, T. 2004. Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import. *Traffic* 5: 847-854.

Ma, H., Tu, L., Naseri, A., Huisman, M., Zhang, S., Grunwald, D. & Pederson, T. 2016. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature Biotechnology* 34: 528-530.

Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335: 716-719.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* 1: 7.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D. & Rebar, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* 29: 143-148.

Mojica, F., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* 60: 174-182.

Moon, S. B., Lee, J. M., Kang, J. G., Lee, N., Ha, D., Kim, D. Y., Kim, S. H., Yoo, K., Kim, D., Ko, J. & Kim, Y. 2018. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3 '-overhang. *Nature Communications* 9: 3651.

Moore, J. & Haber, J. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology* 16: 2164-2173.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashitna, H., Hata, K., Nakashima, K. & Hatada, I. 2016. Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nature Biotechnology* 34: 1060-1065.

Moscou, M. J. & Bogdanove, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* 326: 1501-1501.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. & Kamoun, S. 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology* 31: 691-693.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki, M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., Zhang, F. & Nureki, O. 2018. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361: 1259-1262.

Pfeiffer, P., Thode, S., Hancke, J. & Vielmetter, W. 1994. Mechanisms of overlap formation in nonhomologous DNA end joining. *Molecular and Cellular Biology* 14: 888-895.

Pourcel, C., Salvignol, G. & Vergnaud, G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151: 653-663.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. & Lim, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173-1183.

Ren, B., Liu, L., Li, S., Kuang, Y., Wang, J., Zhang, D., Zhou, X., Lin, H. & Zhou, H. 2019. Cas9-NG greatly expands the targeting scope of the genome-editing toolkit by recognizing NG and other atypical PAMs in rice. *Molecular Plant* 12: 1015-1026.

Sander, J. D., Cade, L., Khayter, C., Reyon, D., Peterson, R. T., Joung, J. K. & Yeh, J. J. 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnology* 29: 697-698.

Schiavinato, M., Marcet-Houben, M., Dohm, J. C., Gabaldon, T. & Himmelbauer, H. 2020. Parental origin of the allotetraploid tobacco *Nicotiana benthamiana*. *Plant Journal* 102: 541-554.

Segal, D., Dreier, B., Beerli, R. & Barbas, C. 1999. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5 '-GNN-3 ' DNA target sequences. *Proceedings of the National Academy of Sciences of the United States of America* 96: 2758-2763.

Shaw, C. H., Leemans, J., Shaw, C. H., Vanmontagu, M. & Schell, J. 1983. A general-method for the transfer of cloned genes to plant-cells. *Gene* 23: 315-330.

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K. & Cigan, A. M. 2016. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nature Communications* 7: 13274.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, D. F. & Zhang, Y. 2016. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Molecular Plant* 9: 1088-1091.

Till, B. J., Burtner, C., Comai, L. & Henikoff, S. 2004. Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research* 32: 2632-2641.

Till, B. J., Colbert, T., Tompa, R., Enns, L. C., Codomo, C. A., Johnson, J. E., Reynolds, S. H., Henikoff, J. G., Greene, E. A., Stein, M. N., Comai, L. & Henikoff, S. 2003. High-throughput TILLING for functional genomics. *Plant Functional Genomics: Methods and Protocols* 236: 205-220.

van Attikum, H., Bundock, P. & Hooykaas, P. 2001. Non-homologous end-joining proteins are required for Agrobacterium T-DNA integration. *Embo Journal* 20: 6550-6558.

Voinnet, O., Rivas, S., Mestre, P. & Baulcombe, D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant Journal* 33: 949-956.

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Hogbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. & de Gier, J. 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America* 105: 14371-14376.

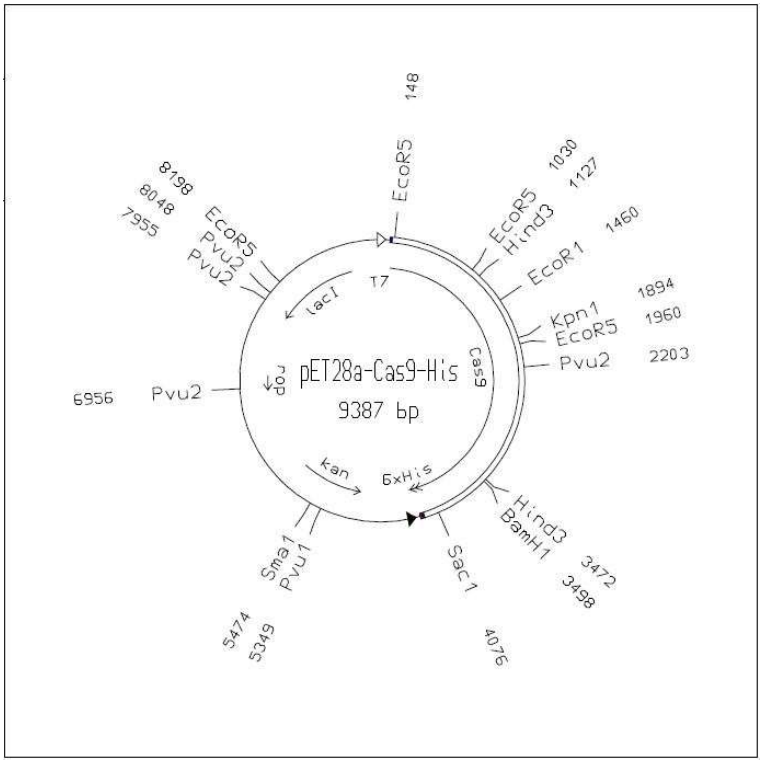
Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. 2020. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368: 290-296.

Woo, J. W., Kim, J., Il Kwon, S., Corvalan, C., Cho, S. W., Kim, H., Kim, S., Kim, S., Choe, S. & Kim, J. 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology* 33: 1162-1164.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V. & Zhang, F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759-771.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J. & Gao, C. 2016. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications* 7: 12617.

APPENDICES



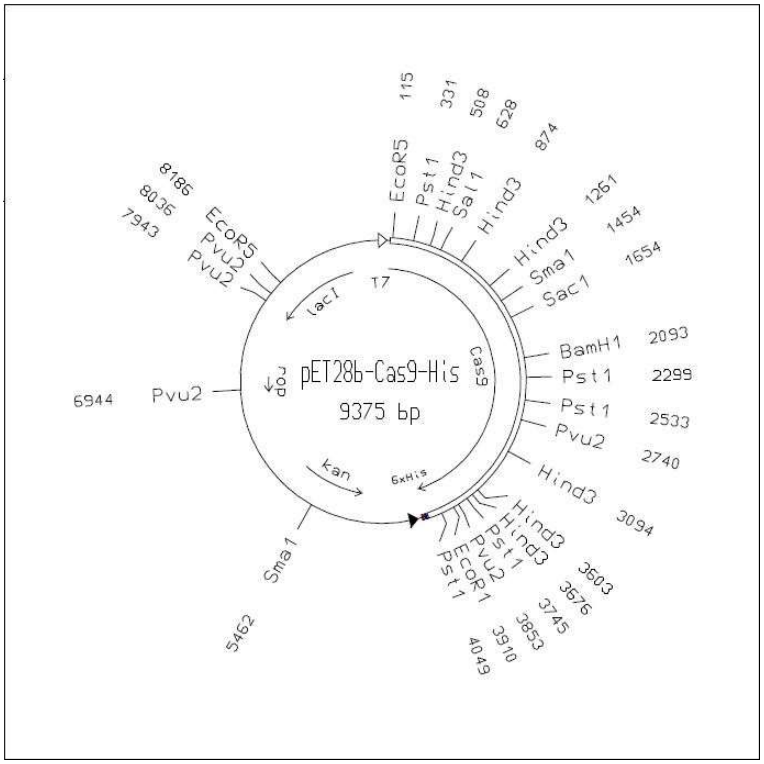


Figure 14. Diagrams of the plasmids pET28a-Cas9-His and pET28b-Cas9-His.

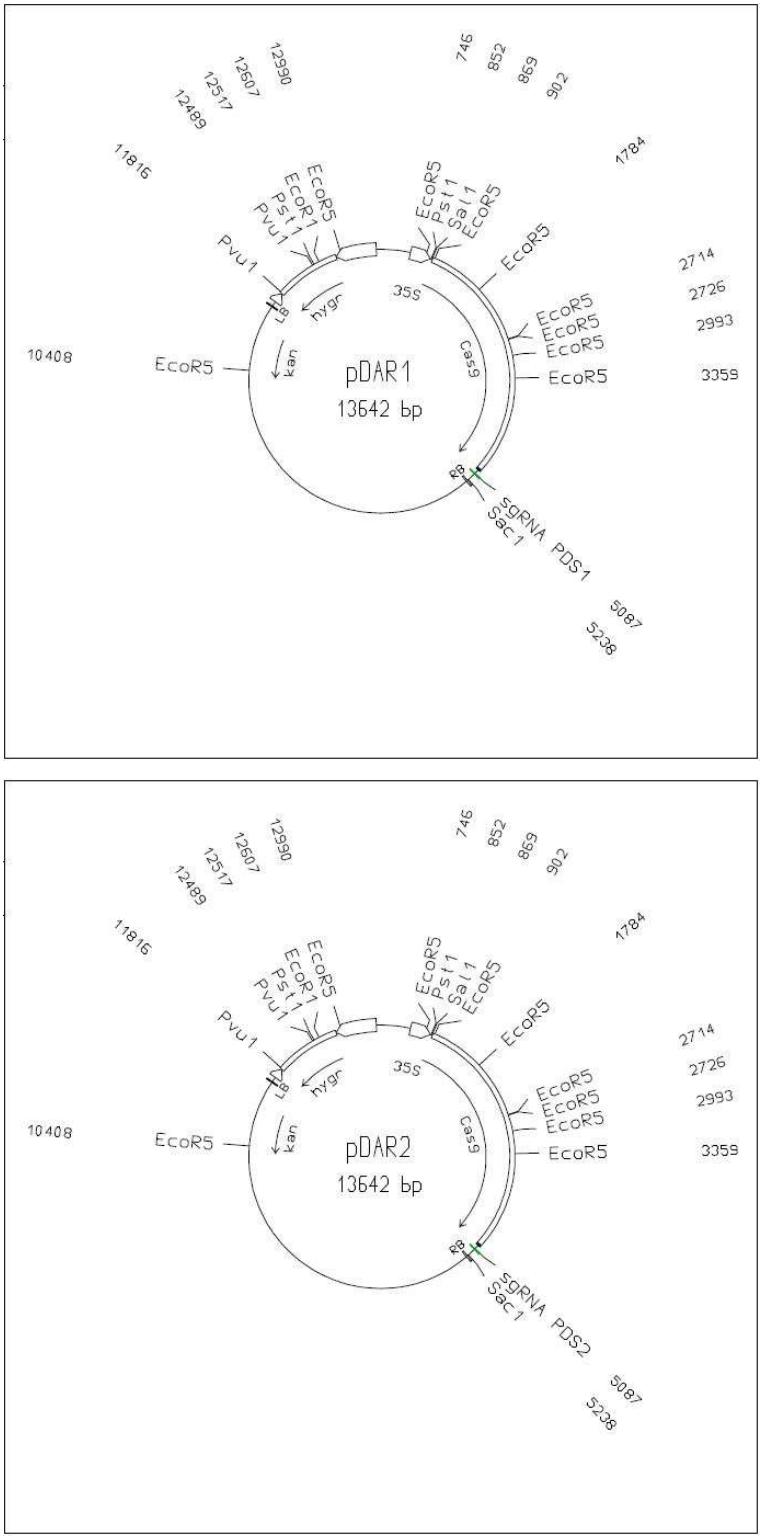


Figure 15. Diagrams of the plasmids pDAR1 and pDAR2 showing hygromycin resistance marker and sgRNAs for *PDS1* and *PDS2*.

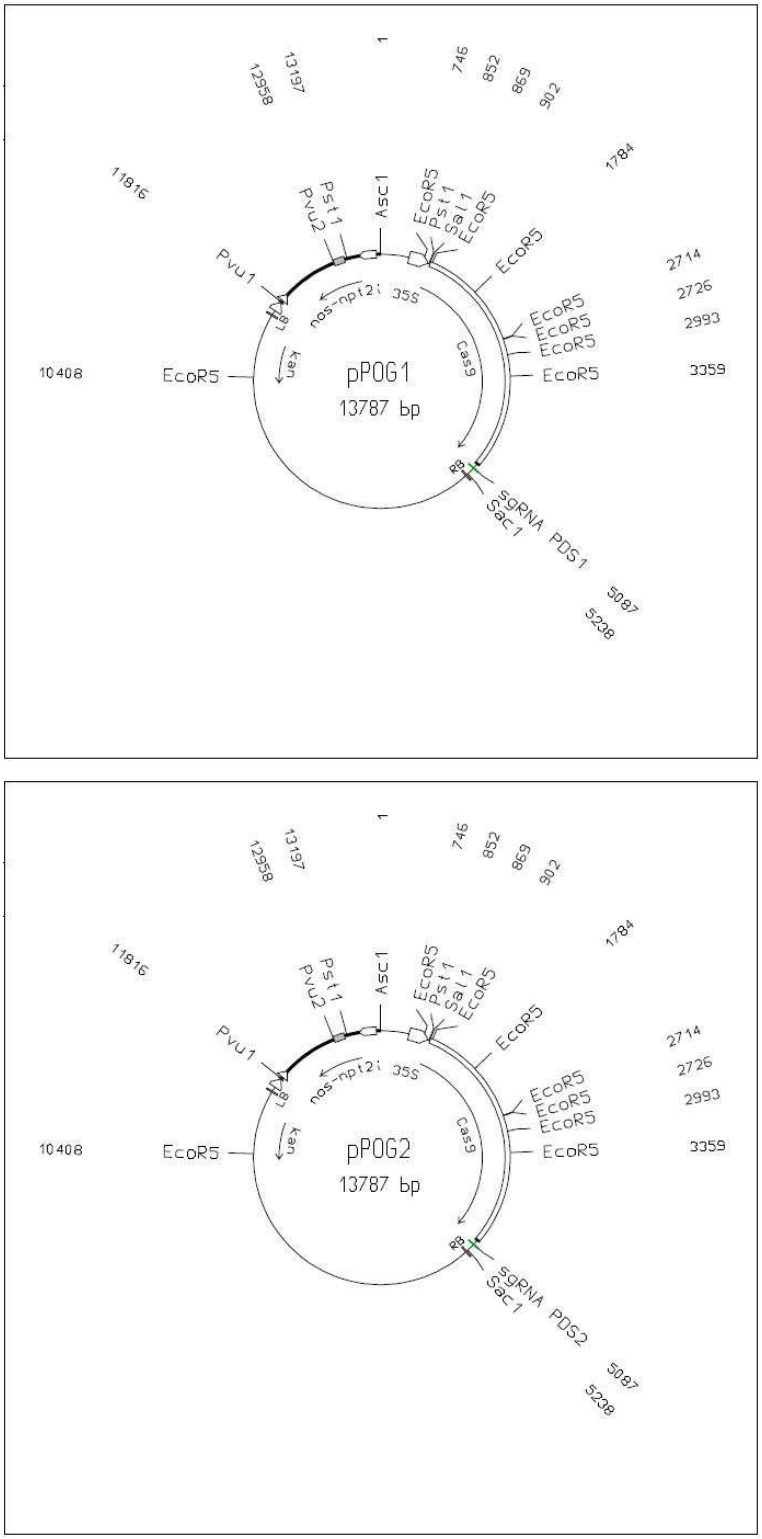
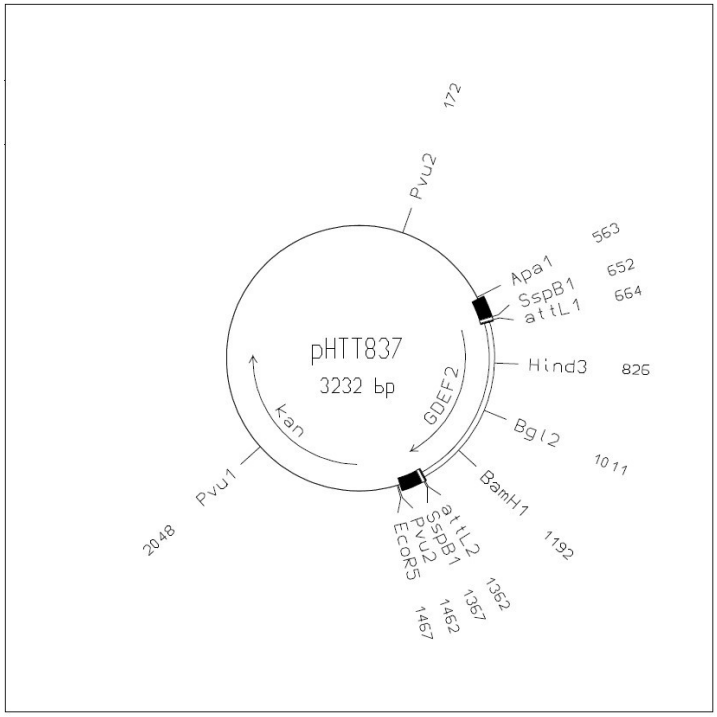
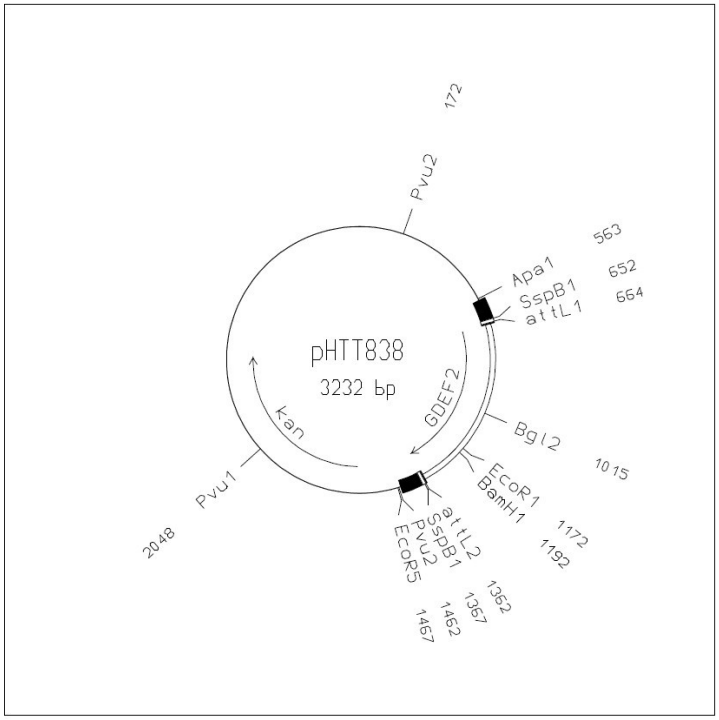


Figure 16. Diagrams of the plasmids pPOG1 and pPOG2 showing added kanamycin resistance marker *Nos-NPTII* and sgRNAs for *PDS1* and *PDS2*.



EcoR1 removed by filling-in in pENTRYgd2



Hind3 removed by filling-in in pENTRYgd2

Figure 17. Diagrams of the plasmids pHTT837 and pHTT838.

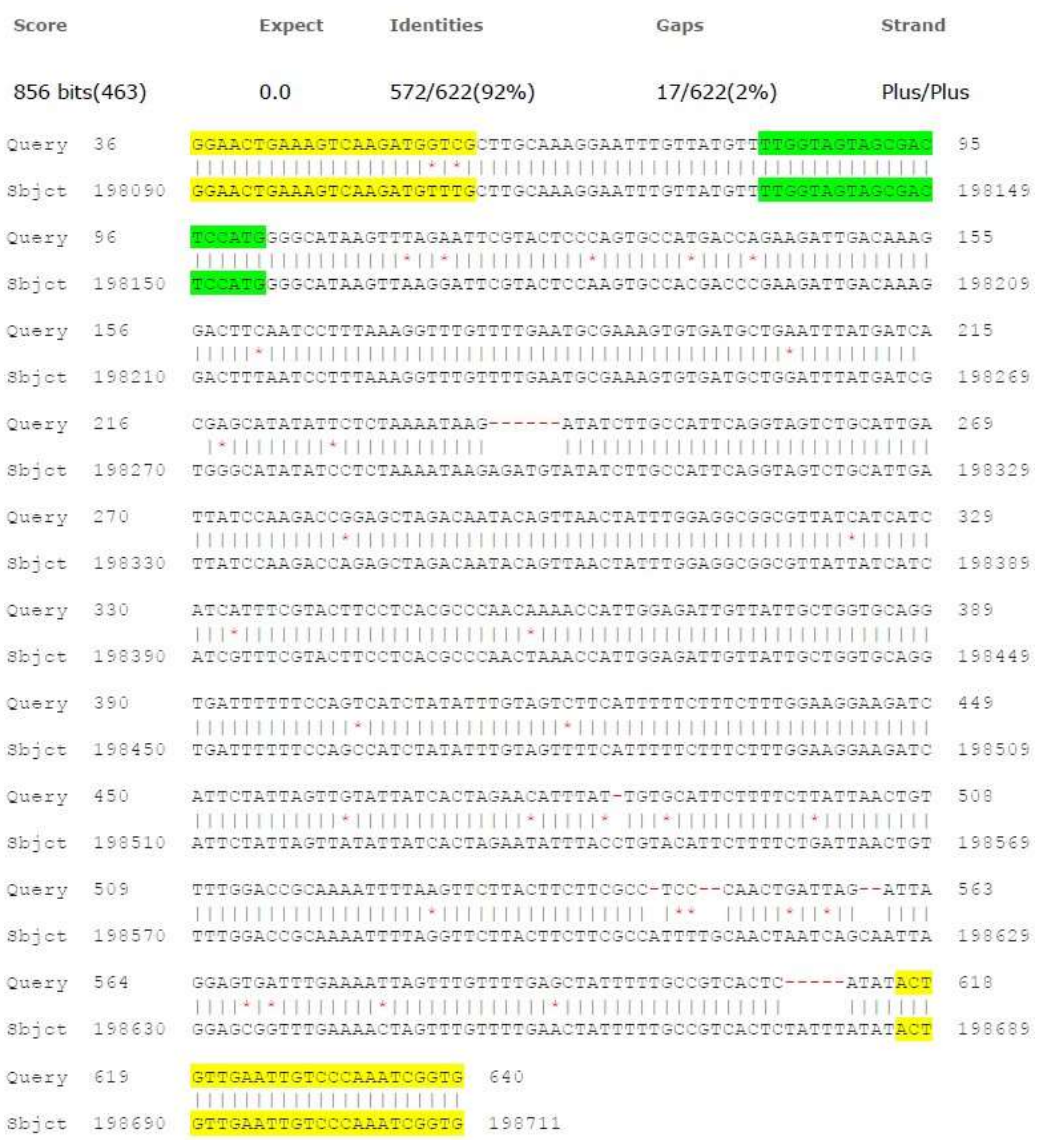


Figure 18. Sequencing results for *N. benthamiana PDS1* fragment in plasmid no. 7 (upper sequence) compared against Niben101Scf01283 genomic sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting.

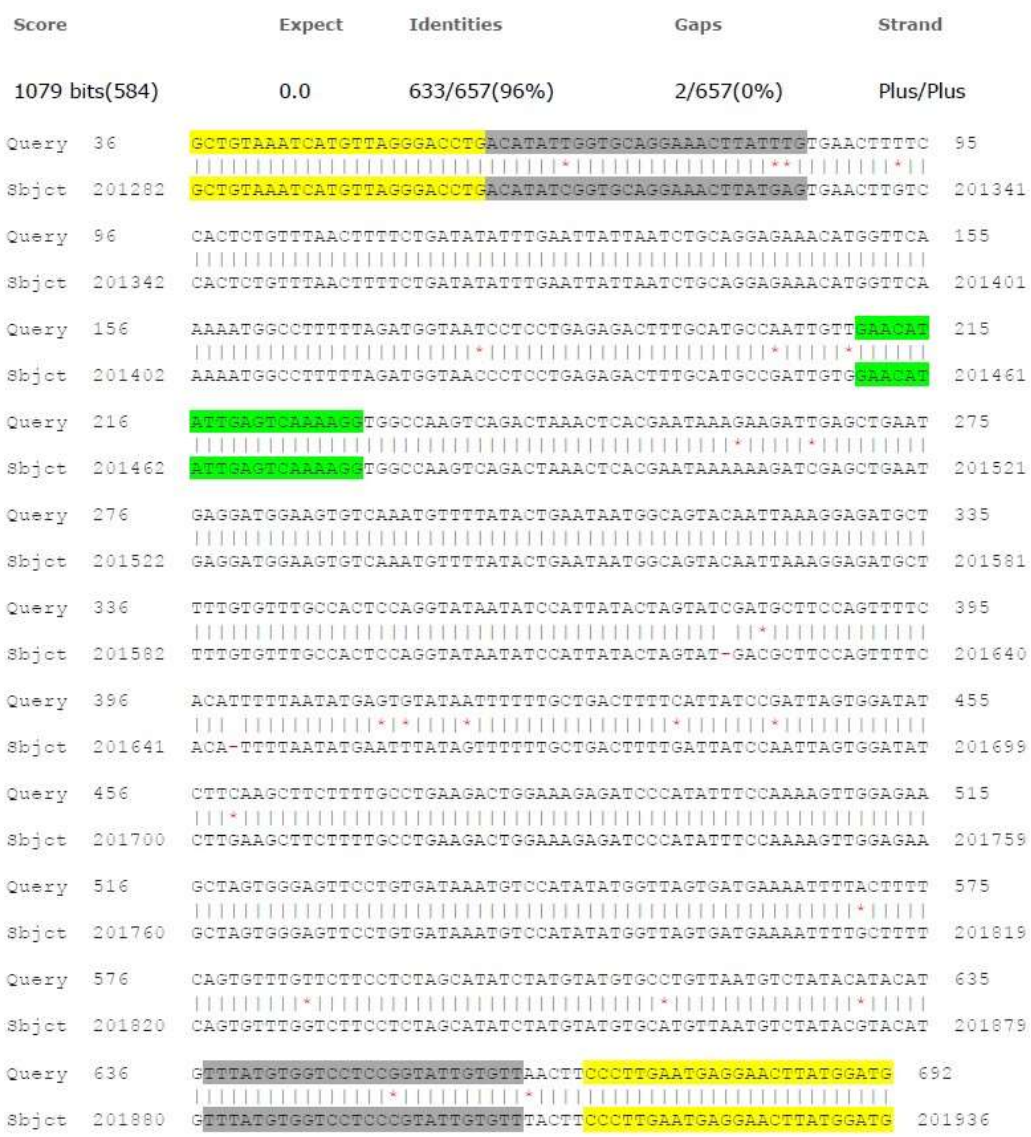


Figure 19. Sequencing results for *N. benthamiana PDS2* fragment in plasmid no. 1 (upper sequence) compared against Niben101Scf01283 genomic sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting. Placement of original non-functional primer pair represented with grey highlighting.

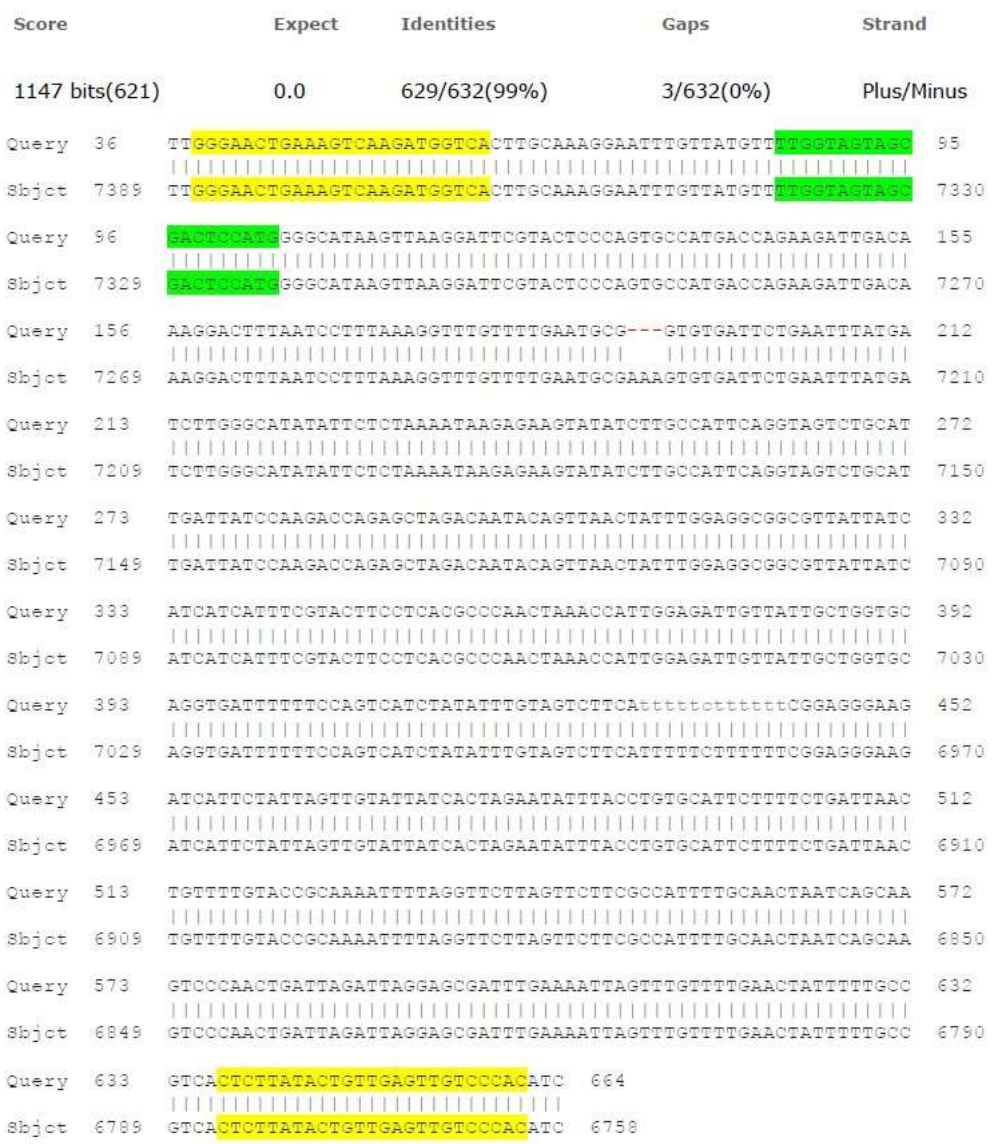


Figure 20. Sequencing results for *N. sylvestris PDS1* fragment from *N. tabacum* in plasmid no. 3 (upper sequence) compared against NW_009521166 genomic *PDS* sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting.

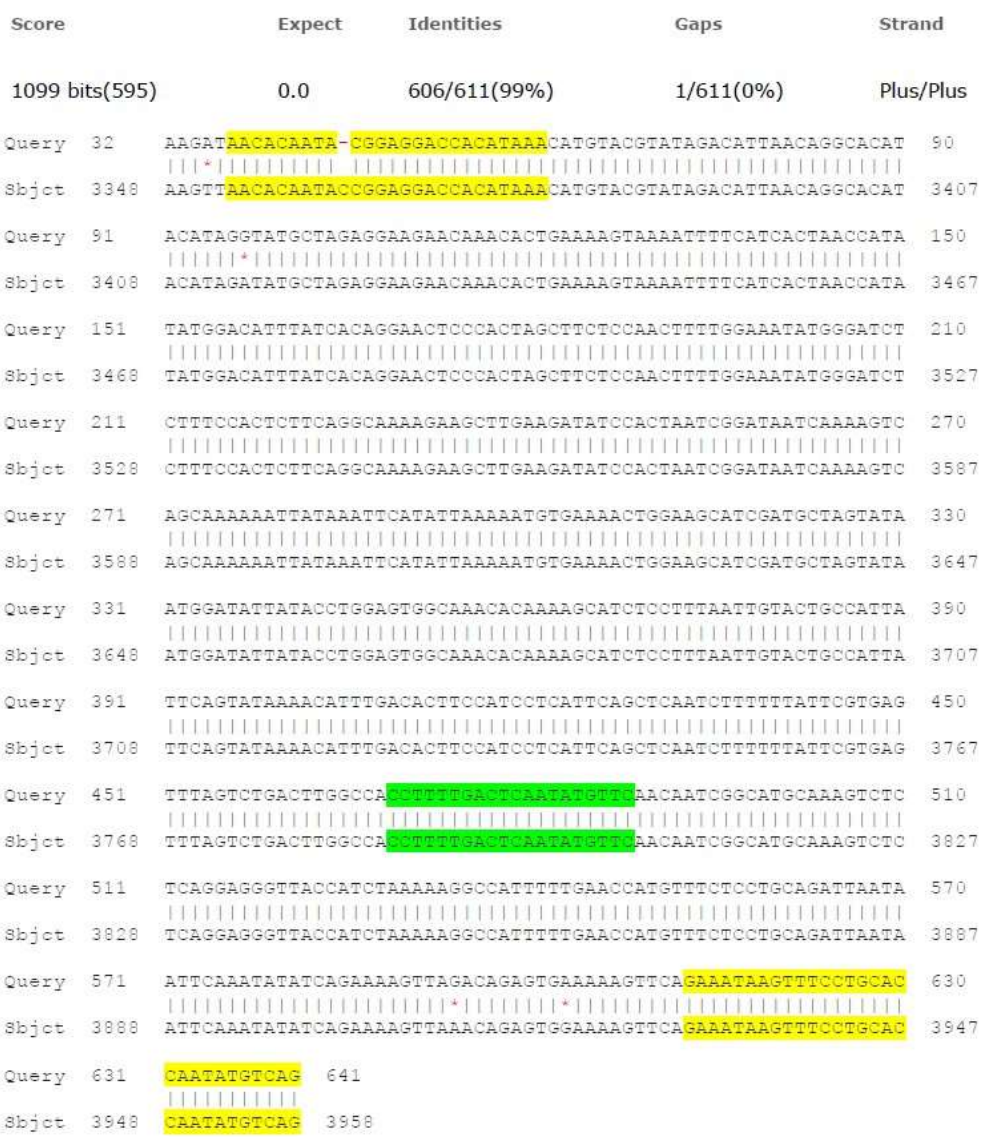


Figure 21. Sequencing results for *N. sylvestris PDS2* fragment from *N. tabacum* in plasmid no. 9 (upper sequence) compared against NW_009521166 genomic *PDS* sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting.

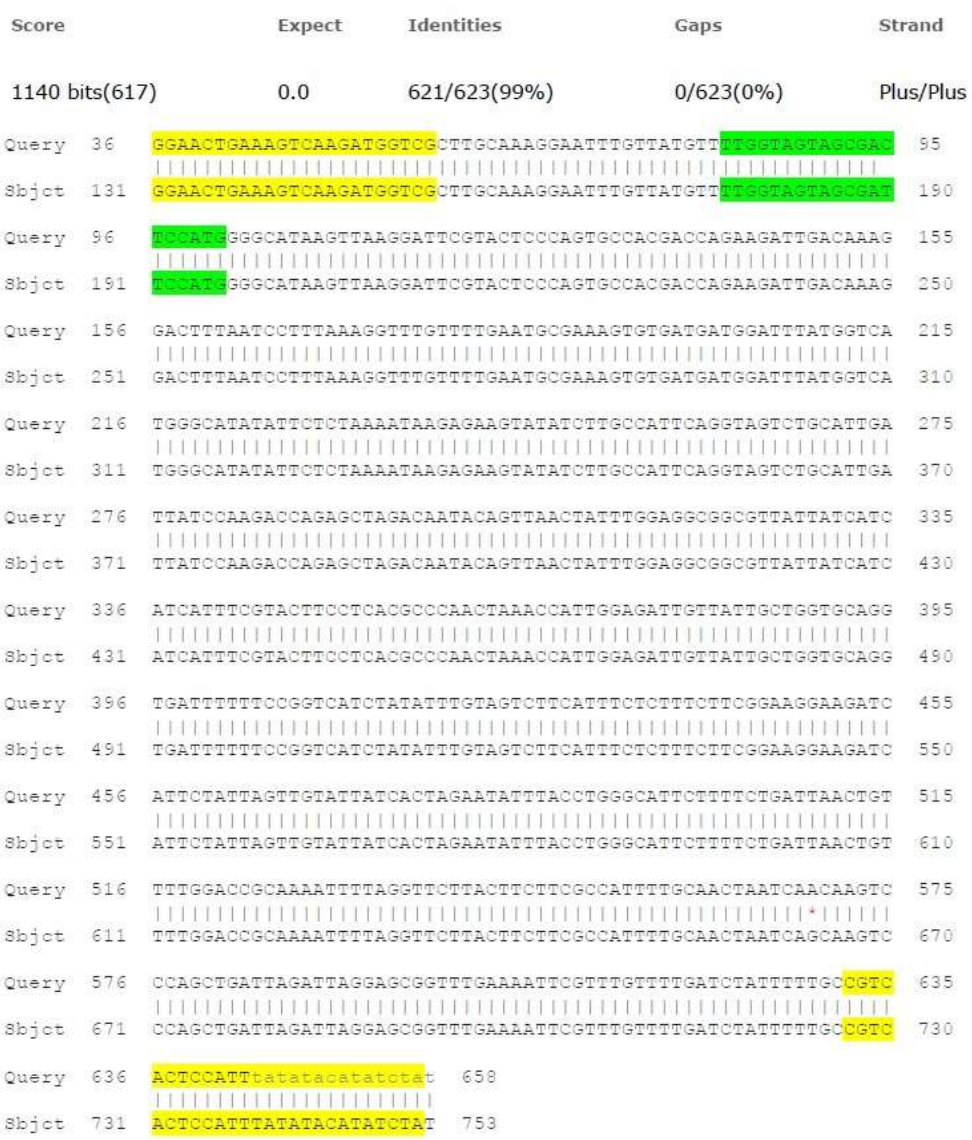


Figure 22. Sequencing results for *N. tomentosiformis PDS1* fragment from *N. tabacum* in plasmid no. 5 (upper sequence) compared against NW_008898264 genomic *PDS* sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting.

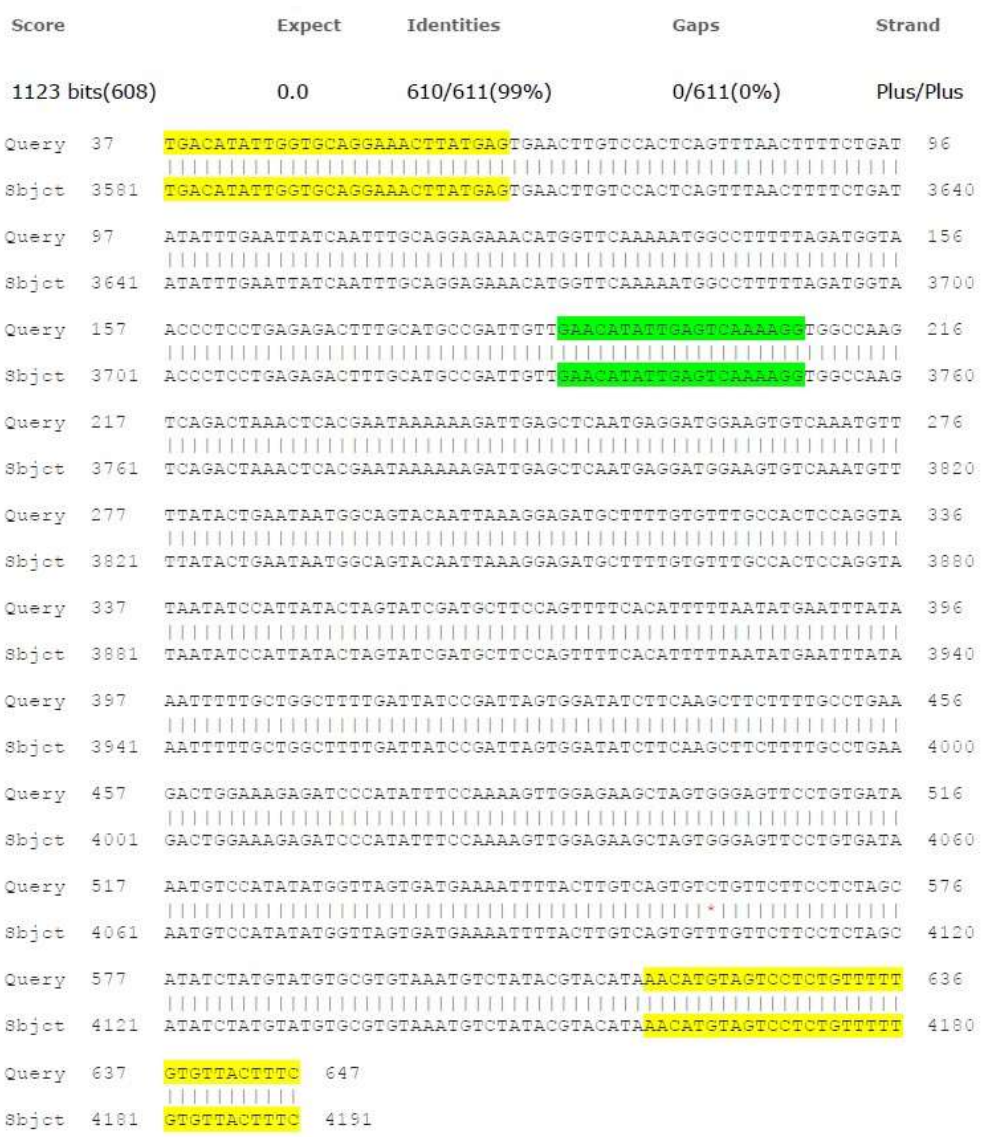


Figure 23. Sequencing results for *N. tomentosiformis PDS2* fragment from *N. tabacum* in plasmid no. 10 (upper sequence) compared against NW_008898264 genomic *PDS* sequence (lower sequence). Primer binding site represented with yellow highlighting. Cas9 target HinfI site represented with green highlighting.